J Forensic Sci, July 2009, Vol. 54, No. 4
doi: 10.1111/j.1556-4029.2009.01080.x
Available online at: www.blackwell-synergy.com

Sree Kanthaswamy,[1,2,3] *Ph.D.; Bradley K. Tom,*[3] *M.S.; Anna-Maria Mattila,*[4] *M.S.; Eric Johnston,*[5] *M.S.; Melody Dayton,*[6] *M.S.; Jennifer Kinaga,*[2] *B.S.; Bethany Joy-Alise Erickson,*[2] *B.S.; Joy Halverson,*[7] *D.V.M.; Dennis Fantin,*[5] *Ph.D.; Sue DeNise,*[5] *Ph.D.; Alexander Kou,*[2] *B.S.; Venkat Malladi,*[2] *B.S.; Jessica Satkoski,*[2] *Ph.D.; Bruce Budowle,*[8,9] *Ph.D.; David Glenn Smith,*[1,2,3] *Ph.D.; and Mikko T. Koskinen,*[4] *Ph.D.*

Canine Population Data Generated from a Multiplex STR Kit for Use in Forensic Casework*

ABSTRACT: Canine biological specimens are often part of the physical evidence from crime scenes. Until now, there have been no validated canine-specific forensic reagent kits available. A multiplex genotyping system, comprising 18 short tandem repeats (STRs) and a sex-linked zinc finger locus for gender determination, was developed for generating population genetic data assessing the weight of canine forensic DNA profiles. Allele frequencies were estimated for 236 pedigreed and 431 mixed breed dogs residing in the U.S. Average random match probability is 1 in 2×10^{33} using the regional database and 1 in 4×10^{39} using the breed dataset. Each pedigreed population was genetically distinct and could be differentiated from the mixed breed dog population but genetic variation was not significantly correlated with geographic transition. Results herein support the use of the allele frequency data with the canine STR multiplex for conveying the significance of identity testing for forensic casework, parentage testing, and breed assignments.

KEYWORDS: forensic science, domestic dog, microsatellites, genotyping, population genetics, database

The use of short tandem repeat (STR) loci for the characterization of human biological evidence is the mainstay of forensic DNA analysis worldwide. STR analysis provides a powerful approach for identity testing, including parentage or kinship analyses. Although not used as widely, the principles and practices for animal DNA identity, parentage and breed analyses or to identify geographical origin or species are the same as those applied in human DNA identity testing. Pet animal DNA analyses have been used in some criminal and civil cases (1–3). Results from STR analyses of biological materials also have been presented in court cases involving livestock such as cattle, goats, horses, llamas, and sheep (http://www.vgl.ucdavis.edu/forensics/index.php).

Dogs are one of the most common domestic pets in the U.S. The American Pet Products Manufacturers Association's (APPMA) National Pet Owners Survey of 2005–2006 estimated that there are over 70 million dogs in the U.S., with an average of 1.7 dogs per household and about one dog per every four persons (2007–2008 APPMA survey). Because of the large number of dogs, it can be expected that (i) canine biological evidence may be found in some criminal cases, (ii) there will continue to be a demand for kinship analysis of expensive breed dogs, and (iii) there will be a need for assessing breed purity. The development and validation of a commercially available canine-specific forensic STR profiling kit would provide such investigations with a valuable tool.

Domestic dog DNA evidence is nominally used in criminal investigations. This situation is due to the unavailability of a standardized and validated canine PCR kit and a lack of standard nomenclature and internal sizing standards (i.e., allelic ladder). The technical inability to obtain meaningful information about the source of canine hair or other biological samples without resorting to specialized laboratories may also contribute to why such evidence is not considered in forensic analyses.

Unlike for human STR typing, there are no commercially available forensic STR multiplex kits for canine DNA testing. Zajc et al. (4), Shutler et al. (5), Padar et al. (6,7), and Halverson and Basten (2) have either used in-house assembled STR panels or kits originally developed for routine parentage testing. These commercial and in-house kits for animal DNA typing were designed for analyzing pristine samples, such as blood or buccal cells which typically contain high quality and quantity DNA. Prior to 2005, the Stockmarks® Canine I and II kits (Applied Biosystems, Foster City, CA) were used effectively to analyze evidentiary canine material in homicide cases (2). These canine kits did not have allelic ladders and required a level of familiarity and scientific expertise not suitable to forensic laboratories. The formalized nomenclature for the

[1]California National Primate Research Center, University of California, Davis, CA 95616.
[2]Department of Anthropology, University of California, Davis, CA 95616.
[3]Graduate Group in Forensic Science, University of California, Davis, CA 95616.
[4]Finnzymes Diagnostics, Keilaranta 16 A, 02150 Espoo, Finland.
[5]MMI Genomics, Inc., 1756 Picasso Avenue, Davis, CA 95618.
[6]The Jan Bashinski DNA Laboratory, California Department of Justice, Richmond, CA 94804.
[7]QuestGen Forensics, 1046 Olive Drive Suite 1, Davis, CA 95616.
[8]Department of Forensic and Investigative Genetics, University of North Texas Health Science Center, Ft. Worth, TX 76107.
[9]Institute of Investigative Genetics, University of North Texas Health Science Center, Ft. Worth, TX 76107.

*This research has been presented at the 2008 NIJ Grantees Conference and at the 2008 International Society for Animal Genetics Conference in Amsterdam, The Netherlands. Findings of this research and parts of this article will be included in the Final Techinal Report to NIJ.

Received 30 June 2008; and in revised form 18 Sept. 2008; accepted 12 Oct. 2008.

Stockmarks® loci was never published and the kits were not updated to the five-dye systems now commonly used in the forensic community. The Stockmarks® Canine kits were discontinued from production in 2005; however, six of the loci in the multiplex described herein were included in the Stockmarks® kits.

A joint initiative has resulted in selection of 18 STR loci for development of a commercial canine DNA profiling kit, the Canine Genotypes™ Panel 2.1 (Finnzymes Oy, Espoo, Finland) for forensic applications. The participants included the California Department of Justice Laboratory in Richmond (CA), MMI Genomics Inc. (MMIG) in Davis (CA), QuestGen Forensics in Davis (CA), the Molecular Anthropology Laboratory (MAL), UC Davis in Davis (CA), the National Institute of Standards and Technology (NIST) in Gaithersburg (MD), the FBI DNA Laboratory in Quantico (VA), the Laboratory of Genomic Diversity, National Cancer Institute in Frederick (MD), and the diagnostics division of Finnzymes Oy in Espoo (Finland).

Canine population genetic databases in the U.S. are predominated by pedigreed dogs (or pure breed dogs who have their ancestry formally recorded by breed registries) and were originally established via routine parentage and pedigree assessment or for other nonforensic objectives. Halverson and Basten (2) included 69 mixed breed dogs (which represent *c.* 50% of the dog population in the U.S., APPMA 2005-06A) out of a total of 558 dogs in their database. In addition to a larger sampling of mixed breed dogs, an ideal canine database would include geographic sampling to assess the distribution of genetic variation across the country.

In accordance with the Scientific Working Group on DNA Analysis Methods, the present report provides a population genetic analysis of 18 polymorphic, canine-specific STRs that can be typed using the kit. A separate developmental *validation* study has been conducted that assessed the robustness and reliability in forensic DNA typing of this multiplex assay which included sensitivity testing, reproducibility studies, intra- and inter-locus color balance studies, annealing temperature and cycle number studies, peak height ratio determination, characterization of artifacts such as stutter percentages and dye blobs, mixture analyses, species-specificity, case type samples analyses (M. Dayton, unpublished data). The loci are displayed in Table 1. This panel of markers, as well as the X and Y chromosome-linked zinc finger loci for gender determination, has been earmarked for formatting into an easy-to-use, quality-controlled system for further development, including the establishment of allelic ladders and a nomenclature system (9), and commercialization.

Population studies were conducted on a sample set that includes geographically distributed populations of mixed breed dogs from the U.S., a sample set that represents dogs from the most popular pedigreed breeds in the U.S. (according to the American Kennel Club [AKC, http://www.akc.org/]), and sample sets of Rottweilers and American Pit Bulls, two dog breeds which have been characterized as "dangerous" by the U.S. Center for Disease Control (CDC) because of the frequency of their involvement in dog bite-related fatalities (18,19). Besides their importance in mauling incidences, the inclusion of Pit Bulls in this study has an additional level of significance. The Pit Bulls are the preferred breed in dog-fighting circles and are thought to be more outbred because of frequent interbreeding with breed types that exhibit the desired phenotypes of aggressiveness and morphology. As such, both the Pit Bull and Rottweiler sample sets represent excellent models for determining the efficacy of our proposed markers in breed and individual identification.

The resulting population database contains publicly accessible information including data on locus informativeness, allele frequencies, distribution of domestic dog genetic variation, match probability estimates, and inbreeding coefficients.

Materials and Methods

DNAs used in this study are from MMIG's collection which includes samples from pedigreed dogs registered with the United Kennel Club (http://www.ukcdogs.com/WebSite.nsf/WebPages/Home) and were extracted from cheek swabs using methods described by DeNise et al. (20).

The PCR was performed in 20 μL volumes containing 2 μL (1.0 ng/μL) template DNA, 9 μL Master Mix (Finnzymes Oy), and 9 μL Primer Mix (Finnzymes Oy) using the AB GeneAmp® PCR System 9700® PCR System (Applied Biosystems). The thermal cycling parameters were: 98°C for 3 min; then 30 cycles of 98°C for 15 sec; 60°C for 75 sec; 72°C for 30 sec, followed by a final 72°C for 5 min. For allele typing, post-PCR amplification products were diluted 1:30 (DNA:high purity water). Two microliters of the diluted amplified product and 0.15 μL of GeneScan-500 [LIZ]® Size Standard (Applied Biosystems) were added to 10 μL of Hi-Di™ formamide (Applied Biosystems), denatured at 95°C for 3 min and followed by snap cooling for 3 min on a StrataCooler Benchtop Cooler (Stratagene, La Jolla, CA).

Electrophoresis was conducted on an ABI PRISM™ 3130-*Avant* Genetic Analyzer (Applied Biosystems) or on an ABI PRISM™ 3100 Genetic Analyzer (Applied Biosystems) in accordance with the instructions in the Finnzymes' Canine Genotypes™ Panel 2.1 manual. The data collection software was set to detect G5 dye chemistry utilizing the DS-33 Dye Primer Matrix Standards Set (Applied Biosystems). Using the Fragment Analysis 36_Pop4 module, the PCR products were injected for 10 sec at 3.0 kV, and subjected to electrophoresis at 15.0 kV at 60°C using the Performance Optimized Polymer (POP™ 7:ABI PRISM™ 3130-*Avant* Genetic Analyzer or POP™ 4:ABI PRISM™ 3100 Genetic Analyzer; Applied Biosystems) in a 36-cm capillary. Fragment sizing was conducted with comparison to a positive control (the Canine Genotypes™ Control DNA001; Finnzymes Oy). GeneMapper® Software v4.0 collection and analysis, ABI PRISM® Data Collection Software v1.1 and GeneScan Analysis v3.7 software packages were used for data collection and size estimation of the fluorescent labeled DNA fragments. ABI PRISM® Genotyper v3.7 NT software was used for automated genotyping of the samples. As an allelic ladder for internal sizing was still under development at the time of the study, the PCR products were binned into their respective allelic categories using the FLEXIBIN program and methods described by Amos et al. (21). The program was modified by the authors to accommodate the repeat motifs of the FH3313 and vWF.X loci (W. Amos, personal communication). The entire data set of raw and binned alleles used in this study is also available (http://www.cstl.nist.gov/biotech/strbase/). Each set of data from the different laboratories was analyzed separately with the FLEXIBIN software. The positive control sample was used to calibrate the allele sizes observed. DNA amplification products from duplicate PCRs of the control animal sample subjected to electrophoresis at different time intervals determined the reproducibility and precision of the assay (see "Run to run sizing"; http://www.cstl.nist.gov/biotech/strbase/).

The chromosomal map coordinates and other relevant information about the 18 autosomal STRs and the sex-linked zinc-finger markers are listed in Table 1. With the exception of VWF.X, a hexameric marker and FH3377, a pentameric marker, all STRs are tetrameric. These include four pairs of syntenic markers: FH2107 and FH3377 on Chromosome 3, FH2054 and PEZ05 on

TABLE 1—*Information on the 18 STR loci and the sex determination locus used in this study.*

Locus	Reference	Repeat Type	Observed Primary Repeat Motif	Estimated Repeat Length	Effective Repeat Range	Chromosome (Map Coordinates*)
FH2001	Francisco et al. (8)	Tetra	GATA	NA	NA	NA
	Tom et al. (9)	Tetra	GATA	4.145	118.77–159.97	23 (50961325–50961475)
FH2004	Francisco et al. (8)	Tetra	GAAA	NA	NA	NA
	Tom et al. (9)	*Tetra*	*AAAG*	*4.197*	*232.82–325.22*	*11 (32161381–32161621)*
FH2010	Francisco et al. (8)	Tetra	ATGA	NA	NA	NA
	Tom et al. (9)	Tetra	ATGA	4.181	221.66–242.66	24 (5196383–5196605)
FH2017	Francisco et al. (8)	Tetra	$GGTA_{(m)}GATA_{(n)}$	NA	NA	NA
	Tom et al. (9)	*Tetra*	$AGGT_{(m)}AGAT_{(n)}GATA_{(o)}$	*3.825*	*256.69–275.69*	*15 (37914470–37914741)*
FH2054	Francisco et al. (8)	Tetra	GATA	NA	NA	NA
	Tom et al. (9)	Tetra	GATA	4.147	139.09–176.53	12 (37914504–37914739)
FH2088	Francisco et al. (8)	Tetra	$TTTA_{(m)}TTCA_{(n)}$	NA	NA	NA
	Tom et al. (9)	Tetra	$TTTA_{(m)}TTCA_{(n)}$	3.971	94.56–138.12	15 (53905651–53905779)
FH2107	Francisco et al. (8)	Tetra	GAAA	NA	NA	NA
	Tom et al. (9)	Tetra	GAAA	3.711	291.72–425.64	3 (83830247–83830574)
FH2309	Ostrander et al. (10)	Tetra	Motif not defined	NA	NA	NA
	Tom et al. (9)	Tetra	GAAA	3.847	339.66–427.98	1 (85772974–85773377)
FH2328	Hellmann et al. (11)	Tetra	GAAA	NA	NA	NA
	Tom et al. (9)	Tetra	GAAA	3.855	171–213.24	33 (19158127–19158477)
FH2361	Mellersh et al. (12)	Tetra	Motif not defined	NA	NA	NA
	Tom et al. (9)	Tetra	GAAA	3.985	322.7–438.7	29 (19723594–19723782)
FH3313	Guyon et al. (13)	Tetra	Motif not defined	NA	NA	NA
	Tom et al. (9)	Tetra	GAAA	3.879	340.93–445.69	19 (24606038–24606459)
FH3377	Guyon et al. (13)	Penta	Motif not defined	NA	NA	NA
	Tom et al. (9)	Penta	GAAAA	4.675	183.01–305.21	3 (78748898–78749090)
PEZ02	Eichmann et al. (14)	Tetra	GGAA	NA	NA	NA
	Tom et al. (9)	Tetra	GGAA	4.011	104.36–144.36	17 (13276076–13276209)
PEZ05	Halverson and Basten (2)/ Halverson et al. (15)	Tetra	AAAG	NA	NA	NA
	Tom et al. (9)	*Tetra*	*TTTA*	*3.967*	*92.48–116.24*	*12 (60326434–60326541)*
PEZ16	Halverson and Basten (2)/ Halverson et al. (15)	Tetra	AAAG	NA	NA	NA
	Tom et al. (9)	*Tetra*	*GAAA*	*3.935*	*280.7–331.66*	*27 (10305692–10305995)*
PEZ17	Halverson and Basten (2)/ Halverson et al. (15)	Tetra	AAAG	NA	NA	NA
	Tom et al. (9)	*Tetra*	*GAAA*	*4.225*	*190.98–224.58*	*4 (71904833–71905038)*
PEZ21	Halverson and Basten (2)/ Halverson et al. (15)	Tetra	AAAT	NA	NA	NA
	Tom et al. (9)	Tetra	AAAT	4.015	83.02–103.22	2 (36438658–36438751)
VWF.X	Shibuya et al. (16)	Hexa	AGGAAT	NA	NA	NA
	Tom et al. (9)	Hexa	AGGAAT	5.965	151.1–186.74	27 (41977918–41978074)
ZFX/ZFY	Aasen and Medrano (17)	–	–	–	–	X/Y

STR, short tandem repeat; NA, information not available.
Italicized fonts indicate a different repeat motif than that previously reported for a particular locus.
*Chromosomal locations were verified using the UCSC Genome Browser (http://genome.ucsc.edu/).

chromosome 12, FH2017 and FH2088 on chromosome 15, and PEZ 16 and vWF.X on chromosome 27.

Only samples with genotypes for all loci were used in our population genetic analysis. The pedigreed dog sample of 236 animals represented nine officially recognized dog breeds, including American Pit Bull Terrier ($N = 38$), Beagle ($N = 34$), Dachshund ($N = 3$), German Shepherd ($N = 35$), Golden Retriever ($N = 32$), Labrador Retriever ($N = 38$), Poodle (Miniature Poodle, $N = 15$; Toy Poodle, $N = 12$; Standard Poodle, $N = 8$), Rottweiler ($N = 15$), Shih Tzu ($N = 4$), and Yorkshire Terrier ($N = 2$). In this study, the Poodles were separated according to their varieties and treated as three separate breeds and therefore, 12 separate pedigreed breeds were actually analyzed. The 431 mixed breed dogs used in this study represent various combinations of 43 different breeds (including Afghan Hound, Akita, American Pit Bulls, Basenji, Basset Hound, Beagle, Belgian Tervuren, Bernese Mountain Dog, Border Collie, Borzoi, Boxer, Bulldog, Chihuahua, Chinese Shar Pei, Chow Chow, Cocker Spaniel, Collie, Dachshund, Doberman Pinscher, English Setter, German Shepherd Dog, German Shorthaired Pointer, Golden Retriever, Greyhound, Italian Greyhound, Labrador Retriever, Mastiff, Miniature Pinscher, Miniature Schnauzer, Mongrel, Poodle, Pug, Rottweiler, Saluki, Samoyed, Scottish Terrier, Shetland Sheepdog, Shih Tzu, Siberian Husky, St. Bernard, Staffordshire Bull Terrier, Whippet, and Yorkshire Terrier) as determined by the Canine Heritage Breed Test™ (22). The samples of pedigreed and mixed breed dogs were acquired from across the U.S. and subdivided into the western ($N = 147$), southern ($N = 241$), mid-western ($N = 164$), and northeastern ($N = 115$) regions for part of the analysis (U.S. Census Bureau [http://www.census.gov/field/www/], see map in Fig. 1).

The exact probability test in GENEPOP version 3.4 (23) was used to test for the presence of linkage disequilibrium (LD, or the nonrandom association of genotypes occurring at different loci) between pairs of the 18 STR loci. To test the null hypothesis that genotypes at one locus segregate independently of genotypes at any other locus at the 0.05 and 0.01 levels of probability, unbiased estimates were made through randomization (1000 iterations) and the Markov-chain method was used to create a contingency table representing the random association of genotypes at all possible pairs of loci. Fisher's method and a sequential Bonferroni-type procedure were used to correct for multiple significance tests (24). In addition

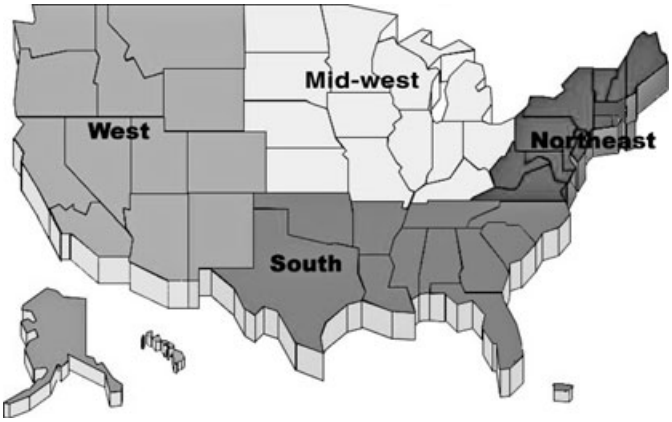

FIG. 1—*Geographic locations in the U.S. that are represented by the samples used in this study.*

to examining LD between loci, Hardy-Weinberg (HW) equilibrium within each locus was analyzed using the GENEPOP software program. Allele frequencies, observed and expected heterozygosities and hierarchical F-statistics and pairwise Fst (25) were computed based on data from all loci using GENEPOP.

To determine if the pedigreed and mixed breed dogs' nuclear genetic variation at the 18 STRs follows some geographic pattern, the STRUCTURE 2.1 software program (26) was used to calculate the expected allele frequencies of individual dogs in each geographic region based on an assignment index and to determine the relative probabilities of assigning each dog to each of the four regions based on the animal's genotypes. Similarly, to distinguish the allele frequencies of the pedigreed and mixed breed populations, STRUCTURE was used to probabilistically assign each dog to a breed category. Both analyses were conducted assuming an admixture model (where animals can represent a mixture of two or more ancestral groups) and correlated allele frequencies among regions and among breeds, respectively. Therefore, when a genotype reflects admixture, or the absence of genetic substructure, a dog will be assigned to two or more populations with probability Q, the proportion of its genome that originated from the Kth population (27). K-values of two to four regions and two to 13 breeds/types (representing mixed breed dogs and the three Poodle types—the Miniature, Toy, and Standard), respectively, were tested so as to include all numbers of possible populations. All STRUCTURE analyses were run at sweeps of 10^4 iterations after a burn-in period of 10^4 with and without *a priori* population information.

The accuracy of assigning individuals to their breed of origin based on genotype data was studied using individual assignment tests, implemented in the program GENECLASS v.2.0 g (28). The program includes several assignment methods but only the Bayesian statistical approach (29) was applied because of its known efficacy (30). Principal component analyses (PCAs) on the regional and breed data sets were also performed using the ADEGENET 1.1 package for R (31).

Results

For the population data set, the effective repeat unit lengths as defined by Amos et al. (21) for the tetrameric loci ranged from 3.7 to 4.2 while the effective repeat unit lengths for pentameric locus FH3377 and the hexameric vWF.X locus were 4.7 and 6.0, respectively. This consistency between observed and estimated repeat unit lengths supported the use of the binning strategy for allele assignment in this study (instead of allelic ladders that are under development). Information on the numbers (n), range, and frequencies for each allele in each of the four U.S. geographic regions and the nationwide sample set (the sum of the four regions) are available (http://www.cstl.nist.gov/biotech/strbase/; see "Observed national and regional STR allele frequencies" [n = number of different allele types]). The number of alleles per locus ranged from 6 to 29 with a mean of 14. A two-way contingency chi-square test for homogeneity of allele frequencies among all four regional populations showed that frequency distributions differed significantly ($p \leq 0.01$) at only three (i.e., FH2010, FH2309, and PEZ05) of the 18 loci studied. Contingency tables for pairs of geographic locations showed that between 0 and 2 loci showed significant allele frequency differences at the $p \leq 0.01$ level. Therefore, allele frequencies did not differ considerably between regions supporting the pooling of regional data and substructure correction concomitant with the pooled data.

The highest pairwise LD was observed in the mid-west and the lowest LD was observed in the northeast (Table 2). In the regional samples, 56 and 36 pairs of loci were significantly associated at the 5% and 1% probability levels, respectively (Table 2). Among the 13 pedigreed and mixed breed dog populations, from 3 (in Beagles) to 19 (in mixed breed dogs) pairs of loci were significantly associated at the 5% level of probability while only one pair of loci (in Beagles, Miniature Poodles, and Rottweilers) and five (in mixed breed dogs) comparisons were significantly associated at the 1% level of probability (Table 3). For each locus pair assayed for the pedigreed dogs, they were statistically significant five times each at the 5% and 1% levels, respectively, while for the mixed breed dogs they were statistically significant 19 and 10 times at the 5% and 1% levels, respectively (Table 3).

Four pairs of physically linked STRs, FH2107 and FH3377, FH2054 and PEZ05, FH2017 and FH2088, and PEZ16 and vWF.X that are located on chromosomes *Cfa* 3, *Cfa* 12, *Cfa* 15, and *Cfa* 27, respectively, did not exhibit greater statistically significant association of genotypes at the $p \leq 0.05$ or $p \leq 0.01$ levels of probability than unlinked loci. Therefore, despite their synteny, these markers were treated similarly to the biologically independent loci and were included in all subsequent population genetic analyses.

Fisher's (32) exact tests across loci within breeds (data not shown) indicate highly significant HW disequilibrium within loci in American Pit Bulls, Golden Retrievers, and the mixed breed dogs. Across breeds, the loci FH2017 and FH2107 were observed to be in highly significant HW disequilibria suggesting nonrandom associations among alleles at these loci only. The nationwide mean observed and expected heterozygosities were 71% and 79%, respectively. Observed heterozygosity at each geographic location ranged from 68% to 73% while the estimated heterozygosity (or gene diversity) ranged from 78% to 79% (Table 4). Across locations, gene diversity estimates were systematically greater than observed numbers of heterozygotes. The data support that while there is a decrease in observed heterozygosity, a high degree of diversity exists within and among breeds and geographic regions.

Table 5 presents breedwise observed and expected heterozygosities whose mean values ranged from >50% (in German Shepherds

TABLE 2—*Pairs of loci in gametic disequilibrium in each of the four geographic regions of the U.S.*

Region	LD at $p \leq 0.05$	LD at $p \leq 0.01$
Western	22	7
Southern	16	6
Mid-Western	59	32
Northeastern	13	13
Across All Regions	56	36

LD, linkage disequilibrium.

TABLE 3—*Pairs of loci in gametic disequilibrium among the 18 STRs in pedigreed and mixed breed dogs.*

Breed	LD at $p \leq 0.05$	LD at $p \leq 0.01$
American Pit Bull	5	2
Beagle	3	1
Dachshund	NA*	NA
German Shepherd*	8	2
Golden Retriever	7	3
Labrador Retriever†	8	5
Miniature Poodle	13	1
Standard Poodle	4	0
Rottweiler	5	1
Shih Tzu	NA	NA
Toy Poodle	NA	NA
Yorkshire Terrier	NA	NA
Across All Pedigreed Dogs	5	5
Mixed Breed Dogs‡	19	10

STR, short tandem repeat; NA, information not available.

*PEZ05 was in LD with FH2054 and FH 3377 in German Shepherds ($p = 0.05$).

†PEZ05 was in LD with FH2010, FH2054, and FH2361 in the Labrador Retrievers ($p = 0.05$).

‡PEZ05 was in LD with FH2054 ($p = 0.05$) and FH2017 was in LD with FH2088 ($p = 0.01$) in the mixed breed dogs.

and Rottweilers) to 69% and 77%, respectively, in Dachshunds. Estimated heterozygosity exceeded observed heterozygosity in all breeds except German Shepherds where these values were approximately equal. When the estimates from all three Poodle breeds were combined, the observed and expected heterozygosities were 64% and 71%, respectively. The Toy Poodles exhibited significantly greater heterozygosity than the combined estimates. Among the mixed breed dogs, the observed and expected heterozygosities were 75% and 79%, respectively.

Fis, Fst, and Fit estimates for each of the regional and national populations are presented in Table 6. The Fis estimate, which measures the degree of inbreeding, was somewhat lower among breeds (0.06) than among regions (0.10) while the degree of genetic differentiation (or genetic subdivision) among the breeds (Fst or fixation index) was much higher (0.09) than that among regions (0.002). Fit, which reflects the combined effects of inbreeding and genetic subdivision, was higher among the dog breeds including mixed breed dogs (0.14) than among the four geographic regions (0.11).

The mean genetic differentiation (pairwise Fst) among populations of dogs from each U.S. region ranged from 0.0006, between the mid-west and the south, to 0.0039, between the west and the northeast (Table 7). The genetic differentiation of the pairwise Fst comparison between breeds ranged from 0.02 (Toy and Standard Poodles) to 0.2788 (Rottweilers and German Shepherds, Table 8).

Tables 4–8 also show no significant differences between estimates of allele frequency, heterozygosity, and inbreeding coefficients based on all linked and unlinked loci and estimates based on only the most informative unlinked loci (i.e., without loci FH2017, FH2107, vWF.X, and PEZ05, each of which had exhibited lower heterozygosity values compared with the locus to which it was physically linked).

Results of the STRUCTURE analyses of the regional samples and each of the 13 pedigreed and mixed breed populations are illustrated in Figs. 2 and 3. These results are concordant with estimates of the allele frequencies, observed and expected heterozygosities and *F*-statistics. Individual assignment tests revealed high assignment success for the purebred dogs with all samples being assigned to their correct reference populations. The average within-breed assignment success score ranged from 99.17% (Poodle) to 100% (Dachshund, Golden Retriever, Labrador Retriever, Rottweiler, Shi Tzu, and Yorkshire Terrier; Table 9). In the mixed breed population, the average assignment score was 97.19% with a range of 38.19–100%. However, 35 of the 431 mixed breed dogs were assigned with a higher likelihood to one of the purebred populations instead of the mixed breed population. Table 9 also presents the complete DNA profile frequency (random match probability, RMP) estimates for breeds with more than 30 animals. Based on fictional breed-specific profiles and without correction for substructure, the RMP was 1.37×10^{-25} for mixed breed dogs compared with 2.47×10^{-20} (German Shepherds) to 6.24×10^{-25} (Beagles).

TABLE 4—*National and regional observed and expected heterozygote frequencies.**

			Regional							
	National (Total)		Western		Southern		Mid-Western		Northeastern	
Loci	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed
FH2001	0.80	0.76	0.77	0.71	0.81	0.78	0.81	0.76	0.80	0.82
FH2004	0.78	0.69	0.79	0.73	0.77	0.69	0.79	0.61	0.80	0.76
FH2010	0.73	0.59	0.72	0.59	0.71	0.57	0.74	0.62	0.73	0.60
FH2017	0.55	0.42	0.51	0.34	0.55	0.45	0.56	0.40	0.58	0.49
FH2054	0.84	0.76	0.85	0.81	0.84	0.76	0.84	0.73	0.84	0.78
FH2088	0.78	0.72	0.79	0.71	0.78	0.73	0.78	0.71	0.76	0.70
FH2107	0.87	0.77	0.86	0.74	0.87	0.75	0.88	0.77	0.89	0.86
FH2309	0.93	0.82	0.92	0.82	0.92	0.84	0.93	0.79	0.93	0.83
FH2328	0.86	0.76	0.85	0.80	0.85	0.78	0.86	0.71	0.87	0.76
FH2361	0.83	0.76	0.81	0.73	0.84	0.77	0.84	0.78	0.84	0.74
FH3313	0.94	0.86	0.93	0.86	0.94	0.88	0.93	0.80	0.93	0.91
FH3377	0.89	0.81	0.88	0.80	0.90	0.86	0.89	0.75	0.88	0.79
VWF.X	0.62	0.56	0.61	0.57	0.63	0.58	0.62	0.52	0.59	0.57
PEZ02	0.77	0.71	0.76	0.70	0.77	0.73	0.78	0.66	0.78	0.74
PEZ05	0.68	0.58	0.67	0.60	0.71	0.60	0.69	0.58	0.63	0.52
PEZ16	0.82	0.73	0.84	0.75	0.83	0.75	0.79	0.65	0.82	0.81
PEZ17	0.80	0.77	0.80	0.76	0.80	0.72	0.81	0.79	0.81	0.86
PEZ21	0.68	0.62	0.67	0.61	0.71	0.64	0.65	0.60	0.67	0.63
Mean	0.79 (0.82)	0.71 (0.74)	0.78 (0.81)	0.70 (0.74)	0.79 (0.82)	0.71 (0.75)	0.79 (0.82)	0.68 (0.71)	0.78 (0.82)	0.73 (0.77)

*Estimates in parentheses are based on only the most informative unlinked loci.

TABLE 5—*Expected (He) and observed (Ho) heterozygote frequencies by breed.**

	He	Ho	He	Ho	He	Ho
Locus	American Pit Bull		Beagle		Dachshund	
FH2001	0.72	0.74	0.79	0.71	0.60	0.67
FH2004	0.68	0.63	0.66	0.53	0.73	0.67
FH2010	0.52	0.53	0.69	0.62	0.60	0.67
FH2017	0.70	0.32	0.47	0.35	0.60	1.00
FH2054	0.77	0.76	0.83	0.76	0.87	0.67
FH2088	0.67	0.66	0.65	0.71	0.87	0.33
FH2107	0.73	0.76	0.85	0.79	0.87	0.67
FH2309	0.82	0.76	0.80	0.74	0.73	0.67
FH2328	0.74	0.71	0.78	0.74	0.93	0.67
FH2361	0.78	0.68	0.86	0.82	0.87	1.00
FH3313	0.89	0.89	0.81	0.79	0.80	0.00
FH3377	0.86	0.79	0.76	0.82	0.87	1.00
VWF.X	0.63	0.63	0.52	0.59	0.60	0.33
PEZ02	0.72	0.61	0.69	0.71	0.53	0.67
PEZ05	0.58	0.45	0.78	0.74	0.80	0.67
PEZ16	0.80	0.79	0.79	0.76	0.93	1.00
PEZ17	0.71	0.68	0.76	0.65	0.80	0.67
PEZ21	0.49	0.47	0.60	0.50	0.80	1.00
Mean	0.71 (0.73)	0.66 (0.69)	0.73 (0.75)	0.68 (0.70)	0.77 (0.78)	0.69 (0.69)
	German Shepherd		Golden Retriever		Labrador Retriever	
FH2001	0.71	0.77	0.74	0.75	0.55	0.55
FH2004	0.49	0.49	0.71	0.84	0.55	0.53
FH2010	0.58	0.57	0.53	0.34	0.31	0.21
FH2017	0.16	0.17	0.62	0.22	0.51	0.53
FH2054	0.72	0.54	0.84	0.84	0.67	0.61
FH2088	0.54	0.46	0.81	0.78	0.67	0.71
FH2107	0.76	0.83	0.82	0.53	0.81	0.68
FH2309	0.61	0.57	0.81	0.88	0.81	0.89
FH2328	0.56	0.51	0.64	0.47	0.54	0.58
FH2361	0.81	0.89	0.72	0.72	0.76	0.71
FH3313	0.71	0.77	0.88	0.88	0.88	0.74
FH3377	0.51	0.54	0.67	0.63	0.86	0.87
VWF.X	0.35	0.37	0.53	0.50	0.49	0.42
PEZ02	0.48	0.46	0.64	0.56	0.77	0.76
PEZ05	0.11	0.11	0.63	0.63	0.71	0.71
PEZ16	0.24	0.26	0.40	0.34	0.81	0.82
PEZ17	0.80	0.91	0.64	0.63	0.80	0.84
PEZ21	0.43	0.43	0.68	0.63	0.71	0.76
Mean	0.53 (0.59)	0.54 (0.58)	0.68 (0.69)	0.62 (0.66)	0.68 (0.69)	0.66 (0.68)
	Miniature Poodle		Standard Poodle		Rottweiler	
FH2001	0.60	0.40	0.77	0.50	0.82	0.80
FH2004	0.76	0.67	0.79	0.63	0.13	0.13
FH2010	0.40	0.40	0.40	0.25	0.54	0.53
FH2017	0.07	0.07	0.44	0.38	0.07	0.07
FH2054	0.82	0.87	0.78	0.50	0.36	0.27
FH2088	0.59	0.67	0.71	0.63	0.44	0.47
FH2107	0.78	0.67	0.85	0.63	0.67	0.73
FH2309	0.82	0.60	0.81	0.75	0.69	0.67
FH2328	0.74	0.53	0.76	0.75	0.64	0.60
FH2361	0.76	0.47	0.59	0.50	0.58	0.60
FH3313	0.69	0.80	0.90	0.75	0.62	0.53
FH3377	0.84	0.73	0.65	0.63	0.78	0.80
VWF.X	0.60	0.67	0.43	0.25	0.69	0.60
PEZ02	0.77	0.93	0.80	0.75	0.76	0.60
PEZ05	0.65	0.40	0.69	0.38	0.43	0.33
PEZ16	0.79	0.73	0.76	0.75	0.74	0.47
PEZ17	0.74	0.73	0.79	0.75	0.73	0.73
PEZ21	0.68	0.67	0.69	0.63	0.48	0.53
Mean	0.67 (0.71)	0.61 (0.66)	0.70 (0.73)	0.58 (0.63)	0.57 (0.59)	0.53 (0.55)
	Shih Tzu		Toy Poodle		Yorkshire Terrier	
FH2001	0.61	0.25	0.73	0.67	0.83	0.50
FH2004	0.86	0.75	0.78	0.92	0.67	0.00
FH2010	0.68	0.25	0.56	0.50	0.67	1.00
FH2017	0.54	0.25	0.29	0.33	0.00	0.00
FH2054	0.75	0.75	0.82	0.75	1.00	1.00
FH2088	0.54	0.25	0.83	0.83	0.83	1.00

TABLE 5—*Continued.*

	He	Ho	He	Ho	He	Ho
Locus	Shih Tzu		Toy Poodle		Yorkshire Terrier	
FH2107	0.86	1.00	0.87	0.75	1.00	1.00
FH2309	0.86	0.75	0.83	0.75	0.83	0.50
FH2328	0.89	0.75	0.79	0.58	1.00	1.00
FH2361	0.79	1.00	0.80	0.75	0.50	0.50
FH3313	0.79	0.50	0.92	0.92	0.50	0.50
FH3377	0.82	0.25	0.82	0.67	0.83	1.00
VWF.X	0.25	0.25	0.75	0.83	0.83	1.00
PEZ02	0.46	0.50	0.82	0.92	0.00	0.00
PEZ05	0.71	0.50	0.74	0.58	0.83	0.50
PEZ16	0.75	0.75	0.82	0.83	0.83	1.00
PEZ17	0.75	1.00	0.76	0.67	1.00	1.00
PEZ21	0.25	0.25	0.69	0.75	0.00	0.00
Mean	0.67 (0.70)	0.56 (0.57)	0.76 (0.78)	0.72 (0.75)	0.68 (0.68)	0.64 (0.64)
	Mixed breeds		Poodle breeds			
FH2001	0.80	0.82	0.70	0.52		
FH2004	0.78	0.74	0.78	0.74		
FH2010	0.72	0.67	0.45	0.38		
FH2017	0.56	0.49	0.27	0.26		
FH2054	0.84	0.81	0.81	0.71		
FH2088	0.78	0.75	0.71	0.71		
FH2107	0.88	0.80	0.83	0.68		
FH2309	0.93	0.87	0.82	0.70		
FH2328	0.86	0.84	0.76	0.62		
FH2361	0.84	0.77	0.72	0.57		
FH3313	0.94	0.90	0.84	0.82		
FH3377	0.90	0.85	0.77	0.68		
VWF.X	0.62	0.58	0.60	0.58		
PEZ02	0.77	0.74	0.80	0.87		
PEZ05	0.68	0.62	0.69	0.45		
PEZ16	0.83	0.79	0.79	0.77		
PEZ17	0.81	0.78	0.76	0.72		
PEZ21	0.69	0.65	0.69	0.68		
Mean	0.79 (0.82)	0.75 (0.78)	0.71 (0.74)	0.64 (0.68)		

*Estimates in parentheses are based on only the most informative unlinked loci.

TABLE 6—*F-statistics among domestic dog populations from the western, southern, mid-western, and northeastern U.S. regions and between pedigreed and mixed breed dogs.*

	Among regions			Between all pedigreed and mixed breed dog groups		
Loci	Fis	Fst	Fit	Fis	Fst	Fit
FH2001	0.0447	0.0036	0.0482	0.0023	0.0768	0.0789
FH2004	0.1192	0	0.1192	0.0567	0.1211	0.171
FH2010	0.1835	0.0059	0.1883	0.095	0.1756	0.2539
FH2017	0.2353	0.0033	0.2378	0.1891	0.1083	0.277
FH2054	0.0904	0.0008	0.0911	0.0621	0.0616	0.1199
FH2088	0.0838	0.003	0.0865	0.0389	0.0961	0.1312
FH2107	0.115	0.001	0.1159	0.1024	0.0399	0.1383
FH2309	0.1092	0.0022	0.1111	0.0659	0.0935	0.1533
FH2328	0.1097	0.0033	0.1127	0.0558	0.1126	0.1621
FH2361	0.0904	0.0001	0.0905	0.0749	0.0404	0.1123
FH3313	0.0835	0.0007	0.0841	0.0529	0.0687	0.1179
FH3377	0.0964	0.0015	0.0978	0.058	0.0841	0.1372
VWF.X	0.089	0.0054	0.0939	0.0565	0.0799	0.1318
PEZ02	0.0842	−0.0002	0.084	0.0508	0.0736	0.1207
PEZ05	0.1495	0.0011	0.1505	0.1029	0.1011	0.1936
PEZ16	0.1091	0.0015	0.1104	0.0572	0.1044	0.1556
PEZ17	0.0433	−0.0008	0.0426	0.0285	0.0356	0.0631
PEZ21	0.0901	0.0022	0.0921	0.0556	0.0782	0.1295
Overall	0.1041 (0.0950)	0.0018 (0.0017)	0.1057 (0.0965)	0.0645 (0.0536)	0.0854 (0.0870)	0.1443 (0.1359)

*Estimates in parentheses are based on only the most informative biologically unlinked loci.

The regional PCA (Fig. 4) reveals no geographic distribution of variation among the domestic dogs in the U.S., while the breed PCA in Fig. 5 demonstrates substantial differentiation among these dog breeds with German Shepherds appearing as outliers.

Discussion

A population study has been carried out using 18 STR loci selected specifically for identity testing of canines. To our

TABLE 7—*Pairwise Fst estimates between domestic dog populations from the four geographic regions in the U.S.**

U.S. region	West	South	Mid-West	Northeast
West	–	0.0021	0.0021	0.0032
South	0.0017	–	0.0010	0.0019
Mid-West	0.002	0.0006	–	0.0002
Northeast	0.0039	0.0026	0.001	–

*Estimates above diagonal are based on only the most informative biologically unlinked loci.

knowledge, this is the first investigation of STR diversity and genetic subdivision among dogs of pedigreed and mixed ancestry across different regions of the U.S.

Regardless of breed, mixed breed, or geographic region, the gene diversity for the combined 18 loci is high. Therefore, it can be anticipated that these loci will be useful for identity testing for most forensic and kinship analyses. However, because of the known selection and inbreeding history of the domestic dog imposed by man, it is important to assess the impact of population substructure and how it may impact estimates of the rarity of a canine STR profile. On average, only three locus pairs per breed across the 12 pedigreed breeds and one population of mixed breed dogs tested were found to be out of linkage equilibrium at the $p = 0.01$ level. The number of pairs of loci out of equilibrium was greater among mixed breed than pedigreed dogs. This observation concurs with that of Halverson and Basten's (2) which was based on private sample collections including samples from the AKC. Even pedigreed German Shepherds, which exhibited the lowest diversity among all dog breeds in this study as well as reported by Parker et al. (33), did not show any substantial LD within and between loci.

Estimates of LD between loci based on the current data set were not statistically significant for any syntenic loci. This could be due to the relative distance between these loci on their respective chromosomes which ranged from 5×10^6 bp to 32×10^6 bp, giving alleles at these loci ample opportunity to segregate independently. While greater per breed sample numbers would be desirable, particularly the sample of two Yorkshire Terriers, three Dachshunds, and four Shih Tzus which may limit generating estimates here for those breeds, the data are consistent with other studies and support that allele frequencies across all loci analyzed here can be used to calculate random match, parentage exclusion, and breed assignment probabilities.

The mean observed and expected heterozygosity estimates reported in the present study especially those for Golden and Labrador Retrievers are higher than comparable measures reported by Irion et al. (34) and DeNise et al. (20). This difference, in combination with the within-locus HW disequilibria among three of the most outbred dog breeds, i.e., the American Pit Bull, Golden Retriever and the mixed breed dogs, imply that inbreeding is not the only factor that shaped the domestic dog genetic structure. While our genetic diversity estimates are comparable with those estimated by Halverson and Basten (2), some cross-study differences are evident. For example, the pedigreed Dachshunds (although a sample of only three animals could have biased our estimates) exhibited the highest degree of genetic diversity among all dogs and Pit Bulls were more genetically diverse than the combined breeds of Poodles in this study. Halverson and Basten (2), however, also ranked the Dachshunds highest in genetic diversity after the Poodles, followed by Yorkshire Terriers, then Pit Bulls. In an 85-breed study by Parker et al. (33), the Dachshunds ranked in the middle in terms of autosomal nuclear diversity; however, this breed has been shown to exhibit the most divergent Y chromosome haplotypes among American, Asian, Australian, and European dog breeds (35). There are several varieties of Dachshunds based on size, coat type, and color, therefore the wide assortment of morphometric differences concur with the high level of genetic diversity within this breed.

While the analysis of regionally diverse samples in our study could have inflated our heterozygosity estimates, Fst measurements show that regional variation contributed only 0.2% of the genetic differences among U.S. dog populations (based on geography). Furthermore, the PCA and STRUCTURE analyses of regional populations are consistent with the outcome that there is little genetic differentiation among groups of mixed-breed dogs originating from different geographical regions within the U.S. Regardless of whether the analysis included *a priori* defined geographic groups, when up to four regional populations of domestic dogs were assumed, i.e., $K = 2$–4, no distinct STR distributions emerged with regard to the four geographic regions.

While the population substructure among regional populations is very small, there is a mild correlation between the amount of genetic distance and geographic distance. Genetic relationships among dog populations are only slightly clinally distributed in the U.S. As examples, dogs from the western and northeastern states are more distantly related compared with dogs from the latter region and the mid-western states.

Regional Fst values are commensurate with extremely low levels of variation among regions and suggest sufficient amounts of gene flow among regions of the U.S. to reduce significant genetic subdivision through genetic drift, as well as STR loci which tend to have relatively high mutation rates. Such low levels of regional variation are consistent with the hypothesis that pet owners take their pets with them when they migrate. Each regional sample also

TABLE 8—*Pairwise Fst estimates between pedigreed breeds and mixed breed dogs.**

Breeds	1	2	3	4	5	6	7	8	9	10	11	12	13
1. American Pit Bull Terrier	–	0.0845	0.1090	0.1935	0.1494	0.1502	0.1445	0.1228	0.1713	0.1005	0.1035	0.0915	0.0505
2. Beagle	0.087	–	0.0681	0.2045	0.1292	0.1135	0.1178	0.0935	0.1670	0.0700	0.0699	0.1105	0.0379
3. Dachshund	0.0993	0.0646	–	0.2331	0.1525	0.1487	0.0594	0.0602	0.1952	0.0412	0.0545	0.1193	0.0347
4. German Shepherd	0.1958	0.2045	0.2394	–	0.2702	0.2486	0.2316	0.2209	0.2818	0.2240	0.1649	0.2722	0.1145
5. Golden Retriever	0.1352	0.1208	0.1465	0.2816	–	0.1734	0.1505	0.1245	0.2190	0.1351	0.9052	0.1479	0.0859
6. Labrador Retriever	0.1429	0.1002	0.1357	0.2547	0.1613	–	0.1691	0.1488	0.1849	0.1377	0.1196	0.1387	0.0641
7. Mini Poodle	0.1385	0.1022	0.0717	0.2268	0.1464	0.1548	–	0.0192	0.2028	0.1464	0.0279	0.1404	0.0695
8. Standard Poodle	0.1199	0.0812	0.0709	0.2346	0.1176	0.1313	0.0221	–	0.1780	0.0945	0.0209	0.0941	0.0474
9. Rottweiler	0.1798	0.1647	0.2081	0.2788	0.2157	0.1895	0.1946	0.1723	–	0.2139	0.2007	0.2217	0.1076
10. Shih Tzu	0.1069	0.0859	0.0498	0.2408	0.1409	0.1399	0.1664	0.1281	0.2474	–	0.0875	0.1381	0.0420
11. Toy Poodle	0.1011	0.0603	0.0589	0.1833	0.0919	0.1056	0.0236	0.0214	0.1929	0.1046	–	0.0886	0.0243
12. Yorkshire Terrier	0.0832	0.0802	0.1011	0.2597	0.126	0.1181	0.1011	0.0574	0.193	0.1415	0.0569	–	0.0549
13. Mixed Breed Dogs	0.0485	0.0357	0.0341	0.1148	0.0812	0.0643	0.0637	0.0492	0.1119	0.0512	0.0245	0.0366	–

*Estimates above diagonal are based on only the most informative biologically unlinked loci.

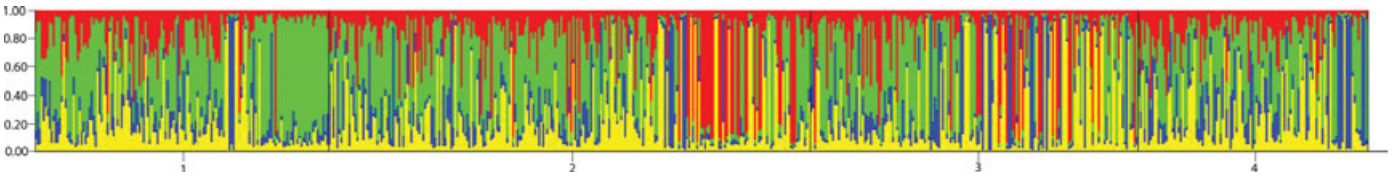

FIG. 2—*Assignment results of the* STRUCTURE *analysis based on the four regional samples (*K = *4). Bar plot of a STRUCTURE analysis using* K = *4, i.e., assuming four distinct regional populations and each represented by different colors. Each individual is represented as a vertical line and is partitioned into* K *colored segments (the four colors) whose length is proportional to the individual's coefficient of membership in the* K *clusters or probability of assignment (*Q*) to the* K*th regional population.*

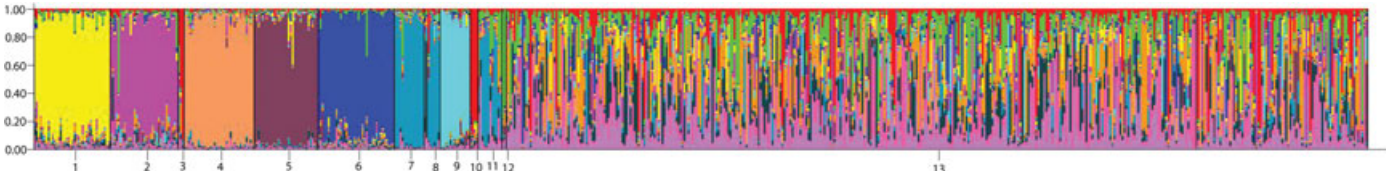

FIG. 3—*Assignment results of the* STRUCTURE *analysis based on the pedigreed breed and mixed breed dog samples, where* K = *13. Bar plot assuming 13 breeds (*K = *13) with each breed represented by a different color. Each individual, represented as a vertical line is partitioned into* K = *13 colored segments whose length is proportional to the individual's probability of assignment (*Q*) to the* K*th breed. (1) American Pit Bull; (2) Beagle; (3) Dachshund; (4) German Shepherd; (5) Golden Retriever; (6) Labrador Retriever; (7) Miniature Poodle; (8) Standard Poodle; (9) Rottweiler; (10) Shih Tzu; (11) Toy Poodle; (12) Yorkshire Terrier; and (13) mixed breed dogs.*

TABLE 9—*Breed assignment probabilities based on* GENECLASS *and RMP estimates for breeds with N > 30.*

Breeds	Range (%)	Average (%)	RMP
American Pit Bull	99.8–100	99.99	2.04×10^{-23}
Beagle	91.9–100	99.74	6.24×10^{-25}
Dachshund	–	100	–
German Shepherd	99.9–100	99.99	2.47×10^{-20}
Golden Retriever	–	100	1.24×10^{-20}
Labrador Retriever	–	100	1.23×10^{-23}
Miniature Poodle	98.1–100	99.84	–
Standard Poodle	93.9–100	99.17	–
Rottweiler	–	100	–
Shih Tzu	–	100	–
Toy Poodle	99.3–100	99.93	–
Yorkshire Terrier	–	100	–
Mixed Breed Dogs	38.2–100	97.19	1.37×10^{-25}

RMP, random match probability.

represented the genetic variation of the much broader nationwide sample, confirming findings of Himmelberger et al. (36) and Baute et al. (37) that were based on canine mitochondrial DNA (mtDNA). As such, in forensic casework, estimates of genetic diversity based on regional populations proximate to the crime scene may also exhibit a similar genetic structure represented by samples collected from different and diverse geographic regions. Conversely, in the absence of a local STR database and no knowledge of the breed or breed make-up of the donor of the biological evidence, investigators can rely on a global database for estimating canine DNA match probabilities.

Previous studies have attributed the wide genetic variation that exists among current dog breeds to each breed's unique breeding history and country of origin (2,34). Among-breed Fst values of 0.09 imply that genetic divergence among domestic dog populations in the U.S. is moderate (although much higher than that for humans—about 10 times the recommended pragmatic value [38]). This among-breed Fst value is comparable with the estimate computed by Halverson and Basten (2). As 9% of variation is attributed to genetic differences among the various dog breeds, *c.* 90% of the genetic diversity is found within breed types. While Fst is influenced by the effective size of and degrees of gene flow among dog populations, most of this variation, especially that among pedigreed dogs and between pedigreed and mixed breed dogs, probably results from strong genetic drift because of small effective population sizes resulting from the artificial selection practices carried out in the domestic dog.

The positive Fis values, in agreement with the gene diversity estimates among breeds and across regions, indicate that there is an increased number of homozygotes in almost all populations of pedigreed dog breeds reflecting the extent of genetic isolation (or inbreeding) still extant among these, if not all, dog breeds. Results based on pedigreed dogs, including the reduced number of loci that are out of equilibrium among pedigreed dogs compared to mixed and pure breed dogs are concordant with the breeding strategies of kennel clubs to outcross pedigreed dogs to maximize their genetic diversity yet maintain the rigid genetic boundaries among breeds to preserve dog breed standards. German Shepherds exhibited more heterozygous individuals than would be expected in an inbred breed, despite being the most highly derived breed, based on low estimates of observed and expected heterozygosity, their genetic divergence from other breeds, and their position in the PCA network.

In the STRUCTURE and the GENECLASS analyses, the dichotomy between pedigreed and mixed breed dogs became obvious, with each of the pedigreed breeds of dogs clustering tightly as distinct genetic groups, and retaining the original assignment probabilities. This is consistent with STR data for humans where the high diversity of the forensically selected loci provides a high power of discrimination, but the populations could be separated according to their known ethnohistory (39). Our STR results are different than those reported by Himmelberger et al. (36) whose mtDNA-based results showed no significant variation between population structures of pedigreed and mixed breed dogs exists. Furthermore, Baute et al. (37) observed no breed-specific 60-bp hypervariable region 1 (HV1) hotspot haplotypes in their study. Even with much larger mtDNA sequences, breed affiliation could not be determined among purported pure bred dogs (36).

The different dog breeds studied here were distributed evenly across the four regions, consistent with AKC's survey (http://www.akc.org/) of top dog breed distributions across U.S. states,

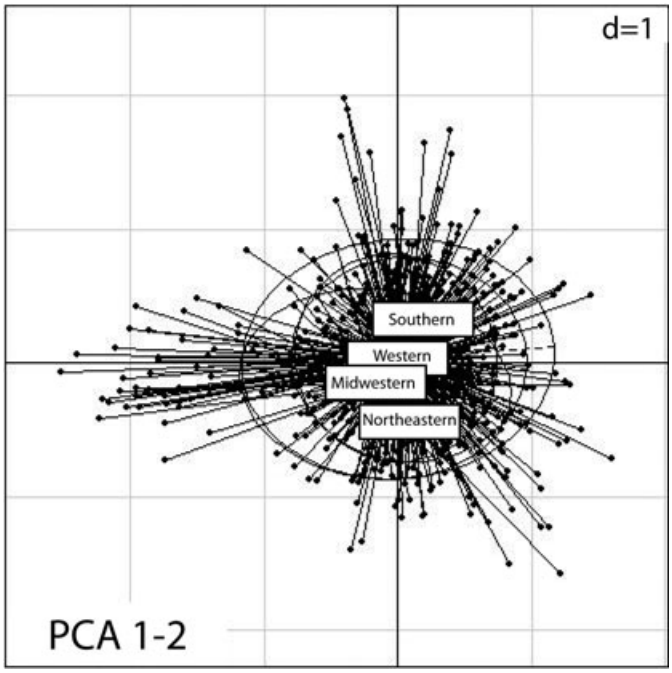

FIG. 4—*Results from the Principal component analyses (PCA) of regional samples.*

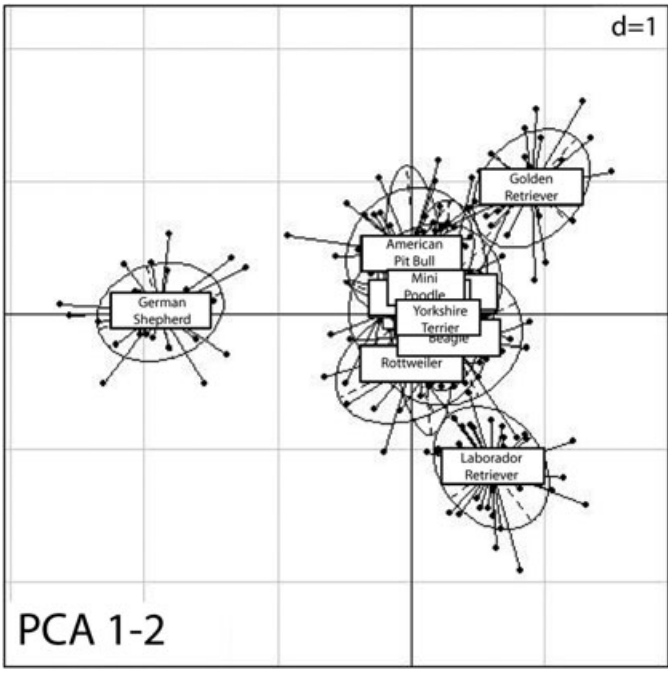

FIG. 5—*Results from the PCA of the different breeds.*

causing regional allele frequency distribution to be very similar even though breed-specific allele frequencies were significantly different. Interestingly, the different Poodle types cluster into a common group despite their morphological differences in size and shape, with Toy Poodles exhibiting the highest degree of genetic introgression from other breeds. Toy Poodles had the most variable gene pool and fewest breed-specific STR alleles than any other breed, including the other poodle-type dogs. The Toy Poodle's genetic composition reflects a much recent but more complex breed development history than the other Poodle subtypes.

The U.S. CDC has attributed most dog-bite fatalities in humans to Pit Bulls and Rottweilers (18,19). Therefore, these two breeds are increasingly involved in litigation to which genetic testing can contribute. We tested the 18 STRs for robustness in identifying these two breeds in particular. While the gene diversity estimates show that the Pit Bulls were more genetically diverse than the combined breeds of Poodles, the STRUCTURE and GENECLASS analyses do not support the notion that Pit Bulls are outbred. Both the pedigreed American Pit Bulls and Rottweilers formed closed clusters in the STRUCTURE analysis, and accordingly their breed assignments, based on the GENECLASS analysis, were 100% successful. This demonstrates that the specific breed categories to which pedigreed Pit Bulls and Rottweilers (as well as the other pedigreed dogs) belong are identifiable based on genetic tests using the 18 STRs reported here. The assignment probability of Pit Bull crosses and Rottweiler crosses were lower, especially for animals assigned to more than two breeds. Furthermore, unlike Himmelberger et al.'s (36) conclusion, our data support that a mixed breed dog can be assigned with some degree of confidence to its predominant breed/s reported by its owners.

As illustrated by the complex patterns of the STRUCTURE analysis, the assignments of mixed breed dogs were indistinct reflecting their mixed ancestry. The mixed breed dogs used in this study represent 43 different breeds and reflect varying degrees of admixture of the breeds. While some of these dogs have been described by their owners and/or have been genetically tested as belonging to a particular breed, they cannot be classified as pedigreed dogs because dog registries require a documented pedigree history and conformation to strict breed standards before recognizing a dog's pedigreed status. Therefore, many of these dogs described as belonging to a specific breed in this study might actually be of mixed ancestry and not breed true-to-type. Also discrepancies in breed definitions could contribute to uncertainty; e.g., in the U.S., the American Pit Bull is sometimes called the Staffordshire Bull Terrier. The rich genetic history of the mixed breed dogs is also evident in their heterozygosity values, which are the highest estimates obtained in this study.

Pairwise comparisons of genetic differentiation (Fst) also demonstrate that mixed breed dogs have a greater degree of genetic similarity with all the other breeds, and therefore reflect a shared ancestry with them. As such, instead of these 18 STR loci to conclusively characterize the heritage of a dog of unknown breed origin, specifically selected breed informative single nucleotide polymorphisms would be needed (33).

Forty-six 608-bp long mtDNA HV1 haplotypes (Smalling unpublished data) including the 16 haplotypes Himmelberger et al. (36) previously reported have been identified among the domesticated dog sample set studied here. A limitation of canine mtDNA haplotyping for identity testing is that most of the sequences can be categorized as common types. Therefore, the mtDNA canine noncoding region tends to have a low discrimination power (compared with that observed in humans). Nonetheless, the genetic marker is still useful for exculpatory purposes and for analyzing questioned samples that are highly degraded (36,37). However, the STR loci studied here appear to be useful to identify the breed of a particular animal that is the subject of litigation, but of greater value is their ability to individualize the DNA of a dog. To evaluate the power of the 18 STRs, a hypothetical canid evidentiary STR profile was used to estimate RMP with Fst correction values of 0.002 and 0.09 for population substructure among the regional populations and among the different breeds, respectively, as recommended by the National Research Council (38).

Following Budowle et al.'s (40) and the National Research Council's (38) recommendations for source attribution, the probability of not observing the evidentiary profile in a population of N unrelated individuals, or $(1 - \text{RMP})N$, should be $\geq (1 - \alpha)$ 100% confidence level. Accordingly, for a confidence level of 99% (where α is 0.01) based on the national allele frequencies, RMP values of less than $1 - (1 - \alpha)1/N = \alpha/N = 1.429 \times 10^{-10}$ are required to have a high degree of confidence that a profile is unique among the estimated $N = 70$ million dogs in the U.S. (40). Estimates of RMP of the fictional profile based on national allele frequencies of 4.8153×10^{-34} with Fst correction of 0.002 for among-region population substructure and of 2.26054×10^{-40} with Fst correction of 0.09 for among-breed population substructure far exceed the threshold for ensuring with a great degree of confidence that the profile is unique among U.S. domestic dogs. Thus, in addition to their ability in identifying the breed composition of an animal that is subject to litigation, the STRs studied here can also identify that animal by its unique genotypic profile with a high level of confidence.

Different breeds yield variations in profile probability estimates because of varying allele frequencies. Without correction for substructure, estimates of RMP are generally more conservative when the breed-specific DNA profile is calculated using the same breed-specific allele frequencies. Weir (41) argued that the use of a suspect's racial data increases the degree of conservativeness. Of course sibs and other first degree relatives may share more genotypes in common and an appropriate conditional kinship analysis should be performed when necessary.

This study focused on the application of the loci included in the proposed canine forensic kit and the associated genetic database in forensic genetic identity and parentage testing in the U.S. The kit's panel of 18 STRs was shown to be informative and robust for identity testing of canines. The database, which is constructed, based on the 18 STRs is more comprehensive than other dog STR databases in terms of regional representation of pedigreed and mixed breed dog populations in the U.S. The genetic profiles and allele frequencies of important dog breeds in the U.S. that are popular as house pets and/or dangerous as vicious animals linked to fatal dog bites are also represented in the database. With enhanced informativity and efficiency as well as their easy accessibility to the forensic laboratories, the kit and the accompanying population genetic database should combine to form a valuable resource that could potentially develop into a universally accepted canine forensic STR system. Lastly, with the availability of a commercial kit, more population data will likely be generated that will enable more precise estimates of the effects of canine population substructure.

Acknowledgments

The authors kindly thank Wayne R. Cavanaugh, President and owner of the United Kennel Club, for his permission to utilize the pedigreed samples analyzed in this study. The authors are grateful to Dr. John Butler, Biotechnology Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, U.S.A., for his contributions to this research. This research was partially funded by the National Institute of Justice (NIJ Grant No. 2004-DN-BX-K007).

References

1. Menotti-Raymond M, Davis V, O'Brian S. Pet cat hair implicates murder suspect. Nature 1997;386:774.
2. Halverson J, Basten C. A PCR multiplex and database for forensic DNA identification of dogs. J Forensic Sci 2005;50(2):1–12.
3. Budowle B, Garofano P, Hellmann A, Ketchum M, Kanthaswamy S, Parsons W, et al. Recommendations for animal DNA forensic identity testing. Int J Legal Med 2005;119(5):295–302.
4. Zajc I, Mellersh C, Kelly EP, Sampson J. A new method of paternity testing for dogs, based on microsatellite sequences. Vet Rec 1994;135(23):545–7.
5. Shutler GG, Gagnon P, Verret G, Kalyn H, Korkosh S, Johnston E, et al. Removal of a PCR inhibitor and resolution of DNA STR types in mixed human-canine stains from a five year old case. J Forensic Sci 1999;44:623–6.
6. Pádár Z, Angyal M, Egyed B, Furedi S, Woller J, Zoldag L, et al. Canine microsatellite polymorphisms as the resolution of an illegal animal death case in a Hungarian zoological gardens. Int J Legal Med 2001;115(2):79–81.
7. Pádár Z, Egyed B, Kontadakis K, Zoldag L, Fekete S. Resolution of parentage in dogs by examination of microsatellites after death of putative sire: case report. Acta Vet Hung 2001;49(3):269–73.
8. Francisco LV, Langston AA, Mellersh CS, Neal CL, Ostrander EA. A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. Mamm Genome 1996;7(5):359–62.
9. Tom BK, Koskinen MT, Dayton M, Maria A-M, Johnston E, Fantin D, et al. Development of a nomenclature system for a canine STR multiplex reagent kit. J Forensic Sci (in press).
10. Ostrander EA, Sprague GF Jr, Rine J. Identification and characterization of dinucleotide repeat (CA)n markers for genetic mapping in dog. Genomics 1993;16:207–13.
11. Hellmann AP, Rohleder U, Eichmann C, Pfeiffer I, Parson W, Schleenbecker U. A proposal for standardization in forensic canine DNA typing: allele nomenclature of six canine-specific STR loci. J Forensic Sci 2006;51:274–81.
12. Mellersh CS, Hitte C, Richman M, Vignaux F, Priat C, Jouquand S, et al. An integrated linkage-radiation hybrid map of the canine genome. Mamm Genome 2000;11(2):120–30.
13. Guyon R, Lorentzen TD, Hitte C, Kim L, Cadieu E, Parker HG, et al. A 1Mb resolution radiation hybrid map of the canine genome. Proc Natl Acad Sci USA 2003;100(9):5296–301.
14. Eichmann C, Berger B, Parson W. A proposed nomenclature for 15 canine-specific polymorphic STR loci for forensic purposes. Int J Legal Med 2004;118:249–66.
15. Halverson J, Dvorak J, Stevenson T, inventors, Metamorphix, Inc., assignee. Microsatellite sequences for canine genotyping. US Patent 5,874,217. 1999 Feb 23.
16. Shibuya H, Collins BK, Huang TH, Johnson GS. A polymorphic (AGGAAT)n tandem repeat in an intron of canine von Willebrand factor gene. Anim Genet 1994;25(2):122.
17. Aasen E, Medrano JF. Amplification of the ZFY and ZFX genes for sex identification in humans, cattle, sheep and goats. Biotechnology 1990;8(12):1279–81.
18. Gershman KA, Sacks JJ, Wright JC. Which dogs bite? Case–control study of risk factors. Pediatrics 1994;93(6):913–7.
19. Sacks JJ, Sinclair L, Gilchrist J, Golab GC, Lockwood R. Breeds of dogs involved in fatal human attacks in the United States between 1979 and 1998. JAVMA 2000;217(6):836–40.
20. DeNise S, Johnston E, Halverson J, Marshall K, Rosenfeld D, McKenna S, et al. Power of exclusion for parentage verification and probability of match of identity in American Kennel Club breeds using 17 canine microsatellite markers. Anim Genet 2004;35(1):14–7.
21. Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS. Automated binning of microsatellite alleles: problems and solutions. Mol Ecol Notes 2007;7(1):10–4.
22. Rosenfeld D, Kerr R, Hutton M, DeNise S, Bates S, Fantin D, inventors. Compositions, methods and systems for inferring canine breeds for genetic traits and verifying parentage of canine animals. Publication number WO/2005/040350. International application number. PCT/US2004/035231. Publication date: 06.05.2005.
23. Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Heredity 1995;86(3):248–9.
24. Rice WR. Analysis tables of statistical tests. Evolution 1989;43(1):223–5.
25. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution 1984;38(6):1358–70.
26. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945–59.

27. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 2003;164:1567–87.
28. Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A. GeneClass2: a software for genetic assignment and first-generation migrant detection. J Hered 2004;95(6):536–9.
29. Rannala B, Mountain JL. Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci USA 1997;94(17):9197–201.
30. Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 1999;153:1989–2000.
31. Jombart T. Adegenet: an R package for the multivariate analysis of genetic markers. Bioinformatics 2008;24(11):1403–5.
32. Fisher RA. Statistical methods for research workers. London: Oliver and Boyd, 1932.
33. Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, et al. Genetic structure of the purebred domestic dog. Science 2004;304(5674):1160–4.
34. Irion DN, Schaffer AL, Famula TR, Eggleston ML, Hughes SS, Pedersen NC. Analysis of genetic variation in 28 dog breed populations with 100 microsatellite markers. J Hered 2003;94(1):81–7.
35. Bannasch DL, Bannasch MJ, Ryun JR, Famula TR, Pedersen NC. Y chromosome haplotype analysis in purebred dogs. Mamm Genome 2005;16(4):273–80.
36. Himmelberger AL, Spear TF, Satkoski JA, George DA, Garnica WT, Malladi VS, et al. Forensic utility of the mitochondrial hypervariable 1 region of domestic dogs, in conjunction with breed and geographic information. J Forensic Sci 2008;53(1):81–9.
37. Baute DT, Satkoski JA, Spear TF, Smith DG, Dayton MR, Malladi VS, et al. Analysis of forensic SNPs in the canine mtDNA HV1 mutational hotspot region. J Forensic Sci 2008;53(6):1325–33.
38. National Research Council. The evaluation of forensic DNA evidence. Washington, DC: National Academy Press, 1996.
39. Budowle B. Population studies on 17 STR loci routinely used in forensic analyses. In: Brinkmann B, Carracedo A, editors. Progress in forensic genetics. Amsterdam: Elsevier, 2003;71–4.
40. Budowle B, Chakraborty R, Carmody G, Monson KL. Source attribution of a forensic DNA profile. Forensic Sci Commun 2000 Available at: http://www.fbi.gov/programs/lab/fsc/backissu/july2000/source.htm.
41. Weir BS. Forensics. In: Balding DJ, Bishop M, Cannings C, editors. Handbook of statistical genetics, 2nd ed. Hoboken, NJ: John Wiley and Sons, 2003;830–52.

Additional information and reprint requests:
Sree Kanthaswamy, Ph.D.
Department of Anthropology
University of California
One Shields Avenue
Davis, CA 95616
E-mail: skanthaswamy@ucdavis.edu