Open Access

Geographic Differences in Mitochondrial DNA (mtDNA) Distribution Among United States (US) Domestic Dog Populations[#]

Briana B. Smalling[1], Jessica A. Satkoski[2], Bradley K. Tom[1], Wing Yi Szeto[2], Bethany Joy-Alise Erickson[2], Theresa F. Spear[1], David Glenn Smith[1,2,3], Bruce Budowle[4], Kristen M. Webb[5], Marc Allard[6] and Sreetharan Kanthaswamy*[,1,2,3]

[1]Graduate Group in Forensic Science, University of California, Davis, CA 95616; [2]Department of Anthropology, University of California, Davis, CA 95616; [3]California National Primate Research Center, University of California, Davis, CA 95616; [4]Department of Forensic and Investigative Genetics, University of North Texas Health Science Center, Fort Worth, TX 76107, USA; [5]Animal Parasitic Diseases Laboratory, Agricultural Research Service, United States Department of Agriculture Building 1180, Beltsville, MD 20705, USA; [6]Molecular Methods and Subtyping Branch, Division of Microbiology, Office of Regulatory Science, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, College Park, MD 20740-3835, USA

Abstract: No geographic differences in mitochondrial DNA (mtDNA) distribution among United States (US) domestic dog populations have been detected to date. To test the hypothesis that regional differences exist, a 608 bp sequence of the canid mtDNA hypervariable region 1 (HV1) from 220 mixed breed animals from the Western, Northeastern, Midwestern, and Southern US were combined with 429 published mixed and pure breed dog HV1 sequences to form a substantial geographically representative dataset. With an increased sample size of regionally representative sequences, geographic substructure among regional populations was shown to be statistically significant using the modified Fisher's exact test and pairwise Fst. The results of the AMOVA showed that 91% of the variation is present within the regional dog populations. Based on these analyses, the significance of regional canine HV1 haplotype distributions and frequencies demonstrate further the value of regional and mixed breed canine mtDNA in forensic investigations in the US.

Keywords: Forensic science, domestic dog, pure breed dogs, hypervariable region 1 (HV1), haplotype, population genetics.

INTRODUCTION

Domestic dogs (*Canis lupus familiaris*) are very popular pets in the United States (US). Thirty-nine percent of US households own at least one dog [1]. Since dog hair can be abundant in an owner's home, it can be readily transferred to and from a crime scene via a variety of intermediate objects such as clothing or vehicles. Thus there is a high likelihood that canine evidence will be present at a crime scene even when little or no human biological evidence may be found.

Biological materials from pure breed or mixed breed dogs have been used to establish important links in human criminal cases such as traffic accidents [2], murders [State of California vs. David Westerfield, 2002], bank robberies [3], and dog attack cases where there are human [4] or non-human victims [5]. In the US alone, there are estimated to be between 3.5 and 4.7 million dog bite injuries to humans annually [6]. In spite of these numbers, canine trace DNA evidence is seldom utilized.

*Address correspondence to this author at the Department of Anthropology, University of California, One Shields Avenue, Davis, CA 95616, USA;
E-mail: skanthaswamy@ucdavis.edu

[#]The research described in this article has been reported in Briana Smalling's Masters of Science (Forensic Science) thesis, which has been submitted to the UC Davis Office of Graduate Studies and has also presented at the 2008 UC Davis Graduate Group in Forensic Science Student Seminar as part of the program's requirement. This research was partially funded by the National Institute of Justice (NIJ Grant No. 2004-DN-BX-K007).

Currently, domestic dog mtDNA diversity studies have been reported for Germany, Sweden, the United Kingdom, China, Japan [7], Portugal [8], and Austria [9]. As a result, databases for forensic use have been compiled in Europe and Asia. Gundry *et al.* [10] compiled one of the largest American domestic dog databases (N=125) which consists of sequences from the entire mtDNA control region. While pure breed dogs have been the focus of these studies as well as more recent studies to create a canine database [10, 11], there have been few attempts within the scientific community to include mixed breed dog samples. In 2004, thirty-nine percent of dogs owned in the US were of mixed breed [1], and mixed breed animals are more common than any specific breed. In order to support the forensic analysis of canine evidence within the US, the present study aims to assess the diversity of canine haplotype frequencies within this canine population and determine the impact of mixed breed domestic dog mtDNA genetic variation on the interpretation of canine mtDNA forensic results. This study will attempt to identify differences within and among the four geographic regions of the US (West, Northeast, Midwest, and South) based on a 608 bp sequence within the mtDNA hypervariable region 1 (HV1).

MATERIALS AND METHODS

For geographic representation, blood samples from domesticated mixed breed dogs were collected from the four regions of the US (West, Northeast, Midwest, and South). To

avoid testing related animals, samples from animals with relatives within the second generation were avoided. DNA extraction and polymerase chain reaction (PCR) amplification was performed using the methods described by Himmelberger *et al*. [12] and Baute *et al*. [13]. For the PCR, the primers H15422 (5'-CTCTTGCTCCACCATCAGC-3') and L16102 (5'-AACTATATGTCCTGAAACCATTG-3'), were used to generate a 720 bp amplified product ranging from nucleotide position (np) 15404 to np 16124 in the canine mtDNA HV1 region [12]. Each PCR contained 2µL of DNA extract, 7.59µL of dH_2O, 0.25µL of dNTPs (10µM), 1.25µL of 10x PCR Rxn Buffer, 0.85µL of $MgCl_2$ (50µM), 0.25µL each of forward and reverse primer (10µM), and 0.06µL of Platinum Taq DNA Polymerase. All PCR reagents ($MgCl_2$, 10x PCR Rxn Buffer, and Platinum Taq DNA Polymerase) were purchased from Invitrogen (Carlsbad, CA). Each set of PCRs included a negative control to monitor potential contamination that may be introduced during PCR preparation. Samples were amplified using the Eppendorf Gradient Mastercycler (Westbury, NY) thermal cycler. The initial hold temperature was 95ºC for 3 minutes followed by 36 cycles of denaturing at 95ºC for 20 seconds, annealing at 55ºC for 10 seconds, and extension at 72ºC for 40 seconds followed by a final hold at 4ºC.

Confirmation of amplification products was performed on 6% polyacrylamide Minigels (BioRad, Hercules, CA), which were submerged in 1x sodium borate (SB) buffer during electrophoresis for approximately 1 hour at 120 volts. Each run contained 2µL of 1x pBR322 DNA-MspI Digest ladder (New England BioLabs, Ipswich, MA) in the first well and 5µL of PCR product plus 1µL of 60% sucrose loading dye in each of the subsequent wells. The PCR products were visualized using a UV light and recorded using an Alpha Innotech imaging system (San Leandro, CA).

Samples were then prepared for sequencing either at the University of California, Davis (UC Davis) DNA Sequencing Facility following the protocol found on their website (http://dnaseq.ucdavis.edu/) or within the UC Davis Molecular Anthropology Laboratory. For in-house sequencing, PCR products were purified with ExoSAP-IT (USB Corporation, Cleveland, OH) according to the manufacturer's protocol. A mixture of 5µL of PCR product and 2µL of ExoSAP-IT were added together and then placed in the thermal cycler for 15 minutes at 37°C, then 15 minutes at 80°C, and followed by a final hold at 4ºC. The BigDye Terminator v3.1 protocol and BigDye XTerminator purification (Applied Biosystems, Foster City, CA) were used for cycle sequencing [14]. The sequence reaction samples were run on a 3130 Genetic Analyzer (Applied Biosystems) using the BigDye Xterminator run module and POP7 polymer (Applied Biosystems).

After cycle sequencing, the forward and reverse sequences from each sample were edited and aligned using Sequencher 4.7 (Gene Codes Corporation, Ann Arbor, MI). These sequences were aligned with the standard reference sequence published by Kim *et al*. [15] (GenBank Accession Number U96639 and labeled as A18 in Table **1** according Pereira *et al*. [11]) to create a 608 bp (np 15456 to np 16063) consensus sequence for each animal.

Fifty-six of the 276 samples used in this study (10 from California, 11 from Colorado, 17 from Massachusetts, 12 from Ohio, and 6 from Texas) could not be sequenced due to either dye blobs, an insufficient amount of PCR product, or poor PCR amplification resulting in high levels of background noise.

Two hundred and twenty 608 bp individual consensus sequences were successfully analyzed and imported into the DNASP program [16] to be aligned with haplotypes from Himmelberger *et al*.'s [12] mixed breed dataset (N = 36) and sequences from Webb *et al*.'s [17] pure breed and mixed breed dataset representing the Western (N=234), Northeastern (N=98) and Southern regions (N=61) (Table **1**). In order to align the sequences properly, two N's were placed at the beginning of each sequence from Webb *et al*.'s [17] dataset since each sequence from this study begins at np 15458. Sequences from Webb *et al*. [17] that had ambiguity codes (N=9) within the 608 bp region were excluded from the analysis leaving sequences with the following GenBank Accession numbers from the Webb *et al*. [17] study to be used in this analysis: EU223385-6, EU223543-6, EU223388-92, EU223548-69, EU223394-8, EU223571-2, EU223400-18, EU223574, EU223420-9, EU223576-601, EU223432-4, EU223603-34, EU223436-43, EU223636-73, EU223445-8, EU223675-6, EU223450-4, EU223678-81, EU223458-70, EU223683-739, EU223472-80, EU223741-61, EU223484-94, EU223763-99, EU223496-509, EU223801-3, EU223511-26, EU223805, EU223528-41, and EU223807-11.

The Webb *et al*. [17] dataset used in this study retained their published names whereas haplotypes unique to this study were named in concordance to Pereira *et al*.'s [11] method (Table **1**) in order to standardize the nomenclature for canine mitochondrial haplotypes in the US. The haplotype fragments generated here and Webb *et al*.'s [17] study varied in length; therefore, three different haplotypes from Webb *et al*. [17] (i.e. B23, B22, and BAmbig8) corresponded to one of the haplotypes generated in this study (Table **1**). Haplotypes unique to this dataset were named in alphanumerical sequence after Webb *et al*.'s [17] haplotypes, i.e., A108-A113, B30-B31, and C12-C13, respectively. Haplotype 8 from Himmelberger *et al*. [12], which corresponds to haplotype C3 in Table **1**, did not align as previously published [12] due to the use of a different alignment program.

No general correlation between breed and mtDNA haplotypes has been established [12, 13, 17]. Therefore, regardless of their breed information, samples from the combined dataset were segregated into four different regions based on the regional divisions defined by the US Census [www.census.gov]: West (comprising samples from California, Colorado, and Nevada), Northeast (comprising samples from Massachusetts and Pennsylvania), Midwest (comprising samples from Ohio), and South (comprising samples from Texas and Virginia). There were no samples from the Midwestern region in Webb *et al*.'s [17] dataset. These regional groupings generated three categories of data: 1) dataset consisting of the entire pure breed and mixed breed domestic dog data (i.e. the combined dataset), 2) only the pure breed data (i.e. the pure breed dataset), and 3) only the mixed breed data (i.e. the mixed breed dataset).

Table 1. The Number (N) and Frequency (*f*) of Each Haplotype (Named According to Pereira *et al.'s* [11] Nomenclature) in the Combined Dataset[1,2]

	West		Northeast		Midwest		South		National	
Haplotypes	N	*f*	N	*f*	N	*f*	N	*f*	N	*f*
A1 (12)	6	0.018	2	0.014	1	0.021	1	0.008	10	0.015
A2	8	0.023	2	0.014	1	0.021	5	0.041	16	0.025
A3	1	0.003	0	0	0	0	0	0	1	0.002
A5	6	0.018	3	0.022	1	0.021	1	0.008	11	0.017
A11 (9)	40	0.117	19	0.138	10	0.208	13	0.107	82	0.126
A14 (14)	1	0.003	0	0	0	0	0	0	1	0.002
A16 (2)	18	0.053	6	0.043	6	0.125	10	0.083	40	0.062
A17 (4)	41	0.12	15	0.109	6	0.125	14	0.116	76	0.117
A18 (1, Ref[‡])	43	0.126	11	0.08	3	0.063	15	0.124	72	0.111
A19	6	0.018	8	0.058	2	0.042	4	0.033	20	0.031
A20	8	0.023	0	0	0	0	0	0	8	0.012
A22 (16)	8	0.023	2	0.014	0	0	1	0.008	11	0.017
A24	3	0.009	0	0	0	0	1	0.008	4	0.006
A26 (3)	6	0.018	2	0.014	0	0	1	0.008	9	0.014
A27	2	0.006	4	0.029	0	0	3	0.025	9	0.014
A28 (5)	3	0.009	1	0.007	0	0	0	0	4	0.006
A29 (6)	5	0.015	1	0.007	1	0.021	2	0.017	9	0.014
A33	0	0	2	0.014	1	0.021	0	0	3	0.005
A40	1	0.003	0	0	0	0	0	0	1	0.002
A66	1	0.003	0	0	0	0	0	0	1	0.002
A68**	1	0.003	2	0.014	0	0	0	0	3	0.005
A70	1	0.003	0	0	0	0	1	0.008	2	0.003
A71**	3	0.009	2	0.014	0	0	1	0.008	6	0.009
A73	0	0	1	0.007	0	0	0	0	1	0.002
A80	2	0.006	2	0.014	0	0	1	0.008	5	0.008
A82	1	0.003	1	0.007	0	0	0	0	2	0.003
A86	1	0.003	2	0.014	2	0.042	1	0.008	6	0.009
A87**	3	0.009	0	0	0	0	0	0	3	0.005
A88	2	0.006	0	0	0	0	2	0.017	4	0.006
A90 (11)	2	0.006	0	0	0	0	0	0	2	0.003
A92	1	0.003	0	0	0	0	0	0	1	0.002
A95	1	0.003	0	0	0	0	0	0	1	0.002
A97	1	0.003	0	0	0	0	0	0	1	0.002
A98	1	0.003	0	0	0	0	0	0	1	0.002

Table 1. Contd….

	West		Northeast		Midwest		South		National	
A101	0	0	0	0	0	0	1	0.008	1	0.002
A102	0	0	1	0.007	0	0	2	0.017	3	0.005
A105	1	0.003	0	0	0	0	0	0	1	0.002
A104	1	0.003	0	0	0	0	0	0	1	0.002
A108 (13)	1	0.003	0	0	1	0.021	0	0	2	0.003
A109 (15)	1	0.003	0	0	0	0	1	0.008	2	0.003
A110	0	0	0	0	0	0	1	0.008	1	0.002
A111	0	0	0	0	0	0	1	0.008	1	0.002
A112	0	0	1	0.007	0	0	0	0	1	0.002
A113	0	0	1	0.007	0	0	0	0	1	0.002
A**	1	0.003	0	0	0	0	0	0	1	0.002
B1 (7)	64	0.187	19	0.138	9	0.188	25	0.207	117	0.18
B3 (10)	5	0.015	1	0.007	0	0	1	0.008	7	0.011
B6	4	0.012	1	0.007	0	0	1	0.008	6	0.009
B8	1	0.003	0	0	0	0	0	0	1	0.002
B10	2	0.006	0	0	0	0	0	0	2	0.003
B11	3	0.009	3	0.022	0	0	0	0	6	0.009
B12	1	0.003	0	0	0	0	0	0	1	0.002
B20	3	0.009	1	0.007	0	0	0	0	4	0.006
B22, B23, BAmbig8**	4	0.012	2	0.014	1	0.021	2	0.017	9	0.014
B25	1	0.003	0	0	0	0	0	0	1	0.002
B28	0	0	0	0	0	0	1	0.008	1	0.002
B29	0	0	1	0.007	0	0	0	0	1	0.002
B30	0	0	1	0.007	0	0	0	0	1	0.002
B31	1	0.003	0	0	0	0	0	0	1	0.002
B1Ambig2	1	0.003	0	0	0	0	0	0	1	0.002
BAmbig11	1	0.003	0	0	0	0	0	0	1	0.002
C1	0	0	1	0.007	1	0.021	0	0	2	0.003
C2	6	0.018	2	0.014	1	0.021	2	0.017	11	0.017
C3 (8)	9	0.026	9	0.065	1	0.021	1	0.008	20	0.031
C8	2	0.006	2	0.014	0	0	4	0.033	8	0.012
C9	0	0	1	0.007	0	0	0	0	1	0.002
C10	1	0.003	1	0.007	0	0	0	0	2	0.003
C12	0	0	1	0.007	0	0	0	0	1	0.002
C13	0	0	0	0	0	0	1	0.008	1	0.002

Table 1. Contd….

	West		Northeast		Midwest		South		National	
CAmbig4	1	0.003	0	0	0	0	0	0	1	0.002
D1	0	0	1	0.007	0	0	0	0	1	0.002
Total	342		138		48		121		649	

[1]The double asterisk (**) denotes ambiguous sequence information in haplotype nomenclature (11) assignments from samples that originated from the Webb et al. (17) dataset.
[1]Haplotypes discovered by Himmelberger *et al.* (12) are in parentheses.
[‡]Ref is the abbreviation for the standard reference sequence published by Kim *et al.* (15) used in this study.

Using ARLEQUIN version 2.001 [18], the Fixation index (Fst) was calculated, and the analysis of molecular variance (AMOVA, [19]) was performed using F statistics to calculate the degree of mtDNA HV1 SNP variation within and among the regional populations of the combined dataset, pure breed dataset, and mixed breed dataset.

The modified Fisher's exact test for homogeneity, as described in Rousset *et al.* [20] and Goudet *et al.* [21], was performed to assess differentiation among regional populations, where the null hypothesis was that haplotype frequencies did not differ among regions. For each of the three datasets described above, pairwise Fst values were generated with 1000 permutations, and average regional population pairwise differences were determined with three calculations: average number of pairwise differences between regional populations (P_{XY}), average number of pairwise differences within the regional populations (P_X and P_Y), and the corrected average pairwise distances between regional populations were calculated as the mean number of pairwise differences between two populations minus the average distance between individuals within those populations or $P_{XY} - (P_X + P_Y)/2$.

The exclusion capacity (probability of exclusion or exclusion power [PE]) for each of the three datasets was calculated using the formula described by Angleby *et al.* [7].

RESULTS

Thirty new 608 bp HV1 haplotypes were discovered among the 220 mixed breed dog sequences (GenBank Accession Numbers FJ501174 – FJ501203) generated in this study. Since none of the Webb *et al.* [17] sequences used in this study are from the Midwest and all of the Midwestern sequences generated in this study are from mixed breed dogs, there are no pure breed data from the Midwestern region. When the combined dataset (Table **1**) was separated into pure breed and mixed breed dog datasets, 21 haplotypes were unique to the mixed breed dataset while 20 haplotypes were unique to the pure breed dataset (Table **2**).

Haplogroups A-D are represented by the mixed and pure breed samples in the combined dataset (Table **1**). Haplotype D1 is from a pure breed Norwegian Elkhound sample from Pennsylvania, and it is the only sample of the purebred Norwegian Elkhound within the combined dataset, which belongs to haplogroup D [17]. Haplotype frequencies varied among regional populations. For example, there is a higher frequency of haplotype A18 in the West (f=0.126) and South (f=0.124) than in the Northeast (f=0.080) and the Midwest (f=0.063) (Table **1**). A significant proportion, (approximately half) of the regional haplotype diversity is contributed by rare haplotypes and some of these are specific to the mixed breed dog fraction of the US canine population. While the frequency of rare haplotypes is similar across regions in the combined dataset (i.e. 25 [f=0.45] rare haplotypes in the West, 17 [f=0.44] in the Northeast, 10 [f=0.59] in the Midwest, 18 [f=0.56] in the South, and 31 [f=0.44] nationally), each regional population has a different composition of rare haplotypes (Table **1**).

There are also different haplotype distributions found in pure breed dogs rather than mixed breed dogs (Table **2**). For instance, 21 out of the 71 haplotypes (29.5%) found in the combined dataset are mixed breed specific whereas 20 out of the 71 haplotypes (28.1%) are pure breed specific. More than half of the mixed breed specific (n=12) and most of the pure breed specific (n=17) haplotypes are rare, i.e., haplotypes that occur only once in the mixed breed and pure breed dogs, respectively. The rare haplotypes of five pure breed dogs (the Finnish Spitz, the Norwegian Elkhound, the Old English Sheepdog, the Swiss Mountain Dog, and the Tibetan Mastiff) are not found in the mixed breed dataset.

The AMOVA analysis for the combined dataset found that the majority of variation is within the regions (91.81% Table **3**). This result was also observed within the pure breed and mixed breed datasets (91.74% and 91.12% respectively Table **3**). When the test for homogeneity was conducted for the combined dataset, significant differences in haplotype frequencies were observed among the different regions at the $p<0.05$ level. The pairwise Fst comparisons for the combined dataset, pure breed dataset, and mixed breed dataset found that all of the observed haplotype distributions from each dataset differ significantly from the random haplotype frequency expectations at the 0.05 level of probability (Table **4**).

Concordant with the fixation indices, the high average numbers of inter- and intrapopulation pairwise Fst differences within the combined, pure breed, and mixed breed datasets reveal that there is substantial genetic differentiation among the different regional populations (Table **4**). Estimates of the PE for each of the combined, pure breed, and mixed breed datasets are presented in Table **5** and are consistent with the previously reported range of the 0.86 - 0.95 [7].

Table 2. The Number (N) and Frequency (f) of the Pure Breed and Mixed Breed Haplotypes (Named According to Pereira *et al.'s* [11] Nomenclature)[1,2]

	Pure Breed								Mixed Breed									
	West		Northeast		South		National		West		Northeast		Midwest		South		National	
Haplotypes	N	f	N	f	N	f	N	f	N	f	N	f	N	f	N	f	N	f
A1 (12)	2	0.011	1	0.014	0	0	3	0.01	4	0.025	1	0.014	1	0.021	1	0.012	7	0.019
A2	7	0.038	1	0.014	3	0.086	11	0.038	1	0.006	1	0.014	1	0.021	2	0.023	5	0.014
A3	0	0	0	0	0	0	0	0	1	0.006	0	0	0	0	0	0	1	0.003
A5	3	0.016	1	0.014	1	0.029	5	0.017	3	0.019	2	0.029	1	0.021	0	0	6	0.017
A11 (9)	14	0.076	13	0.188	3	0.086	30	0.104	26	0.165	6	0.087	10	0.208	10	0.116	52	0.144
A14 (14)	0	0	0	0	0	0	0	0	1	0.006	0	0	0	0	0	0	1	0.003
A16 (2)	6	0.033	2	0.029	1	0.029	9	0.031	12	0.076	4	0.058	6	0.125	9	0.105	31	0.086
A17 (4)	29	0.158	13	0.188	6	0.171	48	0.167	12	0.076	2	0.029	6	0.125	8	0.093	28	0.078
A18 (1, Ref[‡])	24	0.13	5	0.072	2	0.057	31	0.108	19	0.12	6	0.087	3	0.063	13	0.151	41	0.114
A19	3	0.016	2	0.029	0	0	5	0.017	3	0.019	6	0.087	2	0.042	4	0.047	15	0.042
A20	5	0.027	0	0	0	0	5	0.017	3	0.019	0	0	0	0	0	0	3	0.008
A22 (16)	5	0.027	0	0	1	0.029	6	0.021	3	0.019	2	0.029	0	0	0	0	5	0.014
A24	1	0.005	0	0	0	0	1	0.003	2	0.013	0	0	0	0	1	0.012	3	0.008
A26 (3)	5	0.027	0	0	1	0.029	6	0.021	1	0.006	2	0.029	0	0	0	0	3	0.008
A27	1	0.005	1	0.014	2	0.057	4	0.014	1	0.006	3	0.043	0	0	1	0.012	5	0.014
A28 (5)	2	0.011	0	0	0	0	2	0.007	1	0.006	1	0.014	0	0	0	0	2	0.006
A29 (6)	1	0.005	1	0.014	0	0	2	0.007	4	0.025	0	0	1	0.021	2	0.023	7	0.019
A33	0	0	0	0	0	0	0	0	0	0	2	0.029	1	0.021	0	0	3	0.008
A40	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
A66	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
A68**	1	0.005	2	0.029	0	0	3	0.01	0	0	0	0	0	0	0	0	0	0
A70	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	1	0.012	1	0.003
A71**	3	0.016	2	0.029	1	0.029	6	0.021	0	0	0	0	0	0	0	0	0	0
A73	0	0	1	0.014	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
A80	1	0.005	1	0.014	0	0	2	0.007	1	0.006	1	0.014	0	0	1	0.012	3	0.008
A82	0	0	0	0	0	0	0	0	1	0.006	1	0.014	0	0	0	0	2	0.006
A86	0	0	1	0.014	1	0.029	2	0.007	1	0.006	1	0.014	2	0.042	0	0	4	0.011
A87**	3	0.016	0	0	0	0	3	0.01	0	0	0	0	0	0	0	0	0	0
A88	2	0.011	0	0	0	0	2	0.007	0	0	0	0	0	0	2	0.023	2	0.006
A90 (11)	1	0.005	0	0	0	0	1	0.003	1	0.006	0	0	0	0	0	0	1	0.003
A92	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
A95	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0

Table 2. Contd….

	Pure Breed								Mixed Breed											
A97	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
A98	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
A101	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.012	1	0.003
A102	0	0	0	0	0	0	0	0	0	0	1	0.014	0	0	2	0.023	3	0.008
A104	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
A105	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
A108 (13)	0	0	0	0	0	0	0	0	1	0.006	0	0	1	0.021	0	0	2	0.006
A109 (15)	0	0	0	0	0	0	0	0	1	0.006	0	0	0	0	1	0.012	2	0.006
A110	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.012	1	0.003
A111	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.012	1	0.003
A112	0	0	0	0	0	0	0	0	0	0	1	0.014	0	0	0	0	1	0.003
A113	0	0	0	0	0	0	0	0	0	0	1	0.014	0	0	0	0	1	0.003
A**	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
B1 (7)	28	0.152	9	0.13	8	0.229	45	0.156	36	0.228	10	0.145	9	0.188	17	0.198	72	0.199
B3 (10)	3	0.016	1	0.014	1	0.029	5	0.017	2	0.013	0	0	0	0	0	0	2	0.006
B6	2	0.011	1	0.014	1	0.029	4	0.014	2	0.013	0	0	0	0	0	0	2	0.006
B8	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
B10	1	0.005	0	0	0	0	1	0.003	1	0.006	0	0	0	0	0	0	1	0.003
B11	1	0.005	2	0.029	0	0	3	0.01	2	0.013	1	0.014	0	0	0	0	3	0.008
B12	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
B20	1	0.005	1	0.014	0	0	2	0.007	2	0.013	0	0	0	0	0	0	2	0.006
B22, B23, BAmbig8**	2	0.011	2	0.029	1	0.029	5	0.017	2	0.013	0	0	1	0.021	1	0.012	4	0.011
B25	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
B28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.012	1	0.003
B29	0	0	0	0	0	0	0	0	0	0	1	0.014	0	0	0	0	1	0.003
B30	0	0	0	0	0	0	0	0	0	0	1	0.014	0	0	0	0	1	0.003
B31	0	0	0	0	0	0	0	0	1	0.006	0	0	0	0	0	0	1	0.003
B1Ambig2	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
BAmbig11	0	0	0	0	0	0	0	0	1	0.006	0	0	0	0	0	0	1	0.003
C1	0	0	0	0	0	0	0	0	0	0	1	0.014	1	0.021	0	0	2	0.006
C2	4	0.022	0	0	1	0.029	5	0.017	2	0.013	2	0.029	1	0.021	1	0.012	6	0.017
C3 (8)	7	0.038	4	0.058	0	0	11	0.038	2	0.013	5	0.072	1	0.021	1	0.012	9	0.025
C8	1	0.005	0	0	1	0.029	2	0.007	1	0.006	2	0.029	0	0	3	0.035	6	0.017
C9	0	0	1	0.014	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
C10	0	0	0	0	0	0	0	0	1	0.006	1	0.014	0	0	0	0	2	0.006
C12	0	0	0	0	0	0	0	0	0	0	1	0.014	0	0	0	0	1	0.003

Table 2. Contd….

	Pure Breed								Mixed Breed									
C13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.012	1	0.003
CAmbig4	1	0.005	0	0	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
D1	0	0	1	0.014	0	0	1	0.003	0	0	0	0	0	0	0	0	0	0
Total	184		69		35		288		158		69		48		86		361	

[1]The double asterisk (**) denotes ambiguous sequence information in haplotype nomenclature (11) assignments from samples that originated from the Webb et al. (17) dataset
[2]Haplotypes discovered by Himmelberger *et al.* (12) are in parentheses
‡ Ref is the abbreviation for the standard reference sequence published by Kim *et al.* (15) used in this study.

Table 3. MtDNA Fixation Index (Fst) and the Analysis of Molecular Variance (AMOVA) Using the Combined Dataset, Pure Breed Dataset, and Mixed Breed Dataset[1]

	Combined Dataset		Pure Breed Dataset		Mixed Breed Dataset	
Source of mtDNA Variation	D.F.	Variation %	D.F.	Variation %	D.F.	Variation %
Among regions	3	8.19	2	8.26	3	8.88
Within regions	645	91.81	285	91.74	357	91.12
Total	648		287		360	
Fst	0.082		0.083		0.089	

[1]D.F. is the abbreviation for Degrees of Freedom

Table 4. Population Average Pairwise Fst[1,2] and Population Average Pairwise Differences[3] for the Combined Dataset, Pure Breed Dataset, and Mixed Breed Datase[4]

Combined Dataset		West	Northeast	Midwest	South
	West N=342	0.914 (0)	1	1	1
	Northeast N=138	0.075 (0.075*)	0.936 (0)	1	1
	Midwest N=48	0.094 (0.093*)	0.083 (0.082*)	0.897 (0)	1
	South N=121	0.086 (0.086*)	0.075 (0.075*)	0.094 (0.094*)	0.914 (0)
Pure Breed Dataset	West N=184	0.925 (0)	1	--	1
	Northeast N=69	0.083 (0.083*)	0.909 (0)	--	1
	Midwest N=0	--	--	--	--
	South N=35	0.080 (0.080*)	0.088 (0.089*)	--	0.914 (0)
Mixed Breed Dataset	West N=158	0.896 (0)	1	1	1
	Northeast N=69	0.077 (0.078*)	0.951 (0)	1	1
	Midwest N=48	0.103 (0.103*)	0.076 (0.076*)	0.897 (0)	1
	South N=86	0.098 (0.098*)	0.071 (0.071*)	0.097 (0.097*)	0.908 (0)

[1]Parentheses denote the values for the population average pairwise Fst calculation
[2]The asterisk (*) denotes the pairwise Fst comparison that is significantly different from random expectation at the 0.05 level
[3]The values above the diagonal equal the average number of pairwise differences between regional populations. The values on the diagonal equal the average number of pairwise differences within regional populations, and the values below the diagonal equal a corrected average pairwise difference.
[4]The dash (--) denotes where no analysis was performed due to the absence of regional data

Table 5. PEs[1] and Sample Size (N) for the Three Datasets: 1) the Combined Dataset; 2) Pure Breed Dataset; and 3) the Mixed Breed Dataset

Combined Dataset		PE	N
	West	0.912	342
	Northeast	0.929	138
	Midwest	0.878	48
	South	0.907	121
	National	0.916	649
Pure Breed Dataset	West	0.920	184
	Northeast	0.896	69
	South	0.888	35
	National	0.917	288
Mixed Breed Dataset	West	0.890	158
	Northeast	0.937	69
	Midwest	0.878	48
	South	0.898	86
	National	0.908	361

[1]PE is the abbreviation for exclusion power.

DISCUSSION

There is a high level of diversity of mtDNA haplotypes among regional populations of domestic dogs in the US. The frequency distributions of these haplotypes also vary significantly among regions. Based on the regional dataset described herein, if the evidentiary sample was typed as haplotype A18, its lower haplotype frequency in the Midwest than in the West would accord the sample greater evidentiary weight in the Midwest. The regional subdivision described above is also seen in both the pure breed and mixed breed datasets, respectively (Table **2**). This observation is not surprising because dog populations, regardless of breed affiliation, are characterized by male-mediated gene flow and female phylopatry [22] and therefore should exhibit correlations between mtDNA, but not necessarily nuclear DNA, lineages and geographical proximity. As such, some mtDNA haplotypes are common in some geographic locales while being less common or absent in other localities. Correspondingly, a separate study based on a panel of 18 autosomal STRs did not detect any regional differences among dog populations despite using a subset of the canine samples used here [23]. While sporadic associations between pure dog breeds and mtDNA haplotypes are probably an artifact of random genetic drift and selection, the evolution of the domestic dog mtDNA may be too slow compared to that of STRs for meaningful breed assignments [23].

Rare haplotypes contribute disproportionally to regional differentiation among dogs, and since the probability of discovering rare haplotypes depends on sample size, it is likely that this proportion as well as the estimates of inter-regional differences will change as more samples are studied. If a rare haplotype was found in an evidentiary sample recovered from a crime scene, an inclusion would have a relatively higher value than would any of the commonly occurring haplotypes.

A national database would conceal the uniqueness of one region beneath data originating from another region in which the identical haplotype is far more common. For example, haplotype A18 in the combined dataset occurs at a frequency of 0.063 in the Midwest, which is half the frequency of 0.126 and 0.111 in the West and nationally, respectively (Table **1**). Consequently, the PE estimate to weigh the significance of a match involving haplotype A18 based on the national data would be more reflective of the data from the West as opposed to the Midwest. If regional databases are available, an evidentiary sample from the Midwest typed as haplotype A18 would be more forensically valuable than a sample of the same type from the West due to the rarity of haplotype A18 in the Midwest. In order to reduce the possibility of over-weighting the evidence, a regional canine database should be used rather than a national database when a forensic scientist is attempting to assign a significance value to a match.

Previous mtDNA-based studies [12, 13, and 17], where sample sizes were not as large or regionally representative as the dataset used in this study, hence lacking rare haplotypes that are regionally specific, did not detect any regional substructuring among domestic dog populations. With the increase in regional samples (N=649), distinct patterns of regional substructuring have emerged.

The different statistical approaches used in this study and those used by Webb *et al.* [17] may also account for the

disparity in the findings concerning the significance of regional subdivisions. Webb *et al*. [17] relied on the AMOVA analysis of their pure breed, mixed breed, and combined datasets to unravel regional differences; however, based on this study's findings, it is not the AMOVA but rather the Fisher's modified exact test and the pairwise Fst analysis of the various datasets (Table **4**) that revealed the significant subdivisions among the regional populations of the domesticated dog. The differences between the results of both studies could also be attributed to the different sections of the mtDNA that were targeted by both studies. Significantly different Fst values were calculated for the combined, pure breed, and mixed breed datasets (Table **3**) and the correction factor ($p + \theta [1 - p]$, where p is the estimated haplotype frequency and θ is the haplotype-level Fst), derived from Balding *et al*. [24] and described by Budowle *et al*. [25, 26], can be employed on the larger national datasets (Tables **2** and **3**).

If a forensic scientist can distinguish whether the domestic dog sample originated from a pure breed or mixed breed dog, the scientist can use the appropriate Fst value for a more accurate correction of the population substructure within the national database. If not, estimates for both can be calculated. The results from this preliminary study strongly suggest that a database that is more representative of the US regional dog populations and also the mixed breed portion of these dog populations should be compiled for forensic utility.

CONCLUSIONS

Given the continued trend in compiling information on pure bred dogs, mixed breed data are therefore under-represented in current national dog databases. Even though findings by Himmelberger *et al*. [12], Baute *et al*. [13], and Webb *et al*. [17] found that there is comparable genetic diversity among regional and national groupings and a lack of differentiation among them, the hypothesis that regional substructuring among dog populations in the US is supported in this study. The genetic stratification among regional populations, which reflect both diversity within each region as well as differentiation among regions stem from the differences in the frequencies of all haplotypes across regional populations. The average pairwise differences between regional populations among the pure breed and mixed breed datasets, respectively, show that the regional populations vary significantly from each other with the pure breed dogs exhibiting a much higher regional subdivision than mixed breed dogs.

When putting weight on an inclusion, one should take into account the uniqueness of haplotypic distribution and diversity within the region and use a regional database from which the sample originated. Regional dog sampling, whether pure breed or mixed breed, will reflect the diversity within the region and will be more representative of the source of the canine evidence.

ACKNOWLEDGEMENTS

The authors would like to thank the UC Davis Molecular Anthropology Laboratory for the use of their facilities and the UC Davis Graduate Group in Forensic Science for support throughout this project. We would also like to thank reviewers for their critiques and suggestions which helped to improve this manuscript. This project was partially supported by Award No. 2004-DN-BX-K007 awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.

REFERENCES

[1] Economic Analysis of AB1634 Part-2. The source of the dogs: All City Dogs San Diego. http://sandiegodog.wordpress.com. (accessed August 31, **2007**).

[2] Schneider, P.M.; Seo, Y.; Rittner, C. Forensic mtDNA hair analysis excludes a dog from having caused a traffic accident. *Int. J. Legal Med.*, **1999**, *112*, 315-316.

[3] Savolainen, P.; Lundeberg, M.S.; Lundeberg, J. Forensic evidence based on mtDNA from dog and wolf hairs. *J. Forensic Sci.*, **1999**, *44*, 77-81.

[4] Pádár, Z.; Egyed, B.; Kontadakis, K.; Füredi, S.; Woller, J.; Zöldág, L.; Fekete, S. Canine STR analyses in forensic practice. *Int. J. Legal Med.*, **2002**, *116*, 286-288.

[5] Pádár, Z.; Angyal, M.; Egyed, B.; Füredi, S.; Woller, J.; Zöldág, L.; Fekete S. Canine microsatellite polymorphisms as the resolution of an illegal animal death case in a Hungarian Zoological Gardens. *Int. J. Legal Med.*, **2001**, *115*, 79-81.

[6] Eichmann, C.; Berger, B.; Reinhold, M.; Lutz, M.; Parson, W. Canine-specific STR typing of saliva traces on dog bite wounds. *Int. J. Legal Med.*, **2004**, *118*, 337-342.

[7] Angleby, H.; Savolainen, P. Forensic informativity of domestic dog mtDNA control region sequences. *Forensic Sci. Int.*, **2005**, *154*, 99-110.

[8] Von Asch, B.; Pereira, L.; Pereira, F.; Santa-Rita, P.; Lima, M.; Amorim, A. MtDNA diversity among four Portuguese autochthonous dog breeds: a fine-scale characterization. *BMC Genet.*, **2005**, *6*, 37.

[9] Eichmann, C.; Parson, W. Molecular characterization of the canine mitochondrial DNA control region for forensic applications. *Int. J. Legal Med.*, **2007**, *121*, 411-416.

[10] Gundry, R.L.; Allard, M.W.; Moretti, T.R.; Honeycutt, R.L.; Wilson, M.R.; Monson, K.L.; Foran, D.R. Mitochondrial DNA Analysis of the Domestic Dog: Control Region Variation Within and Among Breeds. *J. Forensic Sci.*, **2007**, *52*, 562-572.

[11] Pereira, L.; Von Asch, B.; Amorim, A. Standardisation of nomenclature for dog mtDNA D-loop: a prerequisite for launching a *Canis familiaris* database. *Forensic Sci. Int.*, **2004**, *141*, 99-108.

[12] Himmelberger, A.L.; Spear, T.F.; Satkoski, J.A.; George, D.A.; Garnica W.T.; Malladi, V.S.; Smith, D.G.; Allard, M.W.; Kanthaswamy, S. Forensic utility of the mitochondrial hypervariable 1 region of domestic dogs, in conjunction with breed and geographic information. *J. Forensic Sci.*, **2008**, *53*, 81-9.

[13] Baute, D.T.; Spear, T.F.; Smith, D.G.; Dayton, M.R.; Venkat, M.S.; Vivek, G.; Kou, A.; Kinaga, J.L.; Kanthaswamy, S. Analysis of Forensic SNPs in the Canine mtDNA HV1 Mutational Hotspot Region. *J. Forensic Sci.*, **2008**, *53*, 1325-1333.

[14] Smith, L.M.; Sanders, J.Z.; Kaiser, R.J.; Hughes, P.; Dodd, C.; Connell, C.R.; Heiner, C.; Kent, S.B.; Hood, L.E. Fluorescence detection in automated DNA sequence analysis. *Nature*, **1986**, *321*, 674-679.

[15] Kim, K.S.; Lee, S.E.; Jeong, H.W.; Ha, J.H. The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome. *Mol. Phylogenet. Evol.*, **1998**, *10*, 210-220.

[16] Rozas, J.; Rozas, R. DnaSP version 3: on integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **1999**, *15*, 174-175.

[17] Webb, K.M.; Allard, M. Identification of forensically informative SNPs in the domestic dog mitochondrial control region. *J. Forensic Sci.*, **2009**, *54*, 289-304.

[18] Schneider, S.; Roessli, D.; Excoffier, L. Arlequin: A software for population genetics data analysis. Ver. 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, **2000**.

[19] Excoffier, L.; Smouse, P.; Quattro, J. Analysis of molecular variance inferred from metric distances among DNA haplotypes:

Application to human mitochondrial DNA restriction data. *Genetics*, **1992**, *131*, 479-491.

[20] Rousset, F.; Raymond, M. Testing heterozygote excess and deficiency. *Genetics*, **1995**, *140*, 1413-1419.

[21] Goudet, J.; Raymond, M.; de Meeüs, T.; Rousset, F. Testing differentiation in diploid populations. *Genetics*, **1996**, *144*, 1933-1940.

[22] Parker, H.G.; Kim, L.V.; Sutter, N.B.; Carlson, S.; Lorentzen, T.D.; Malek, T.B.; Johnson, G.S.; DeFrance, H.B.; Ostrander, E.A.; Kruglyak, L. Genetic structure of the purebred domestic dog. *Science*, **2004**, *304*, 1160-1164.

[23] Kanthaswamy, S.; Tom, B.K.; Mattila, A-M.; Johnston, E.; Dayton, M.; Kinaga, J.; Erickson, BJ-A.; Halverson, J.; Fantin, D.; DeNise, S.; Kou, A.; Malladi, V.; Satkoski, J.; Budowle, B.; Smith, D.G.; Koskinen, M.T. Canine population data generated from a multiplex STR kit for use in forensic casework. *J. Forensic Sci*., **2009**, *54*, 829-840.

[24] Balding, D.J.; Nichols, R.A. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci. Int.*, **1994**, *64*, 125-140.

[25] Budowle, B.; Adamowicz, M.; Aranda, X.; Barna, C.; Chakraborty, R.; Eisenberg, A.J.; Frappier, R.; Gross, A.M.; Lee, H.S.; Milne, S.; Prinz, M.; Saldanha, G.; Krenke, B.E. Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. *Forensic Sci. Int.*, **2005**, *150*, 1-15.

[26] Budowle, B.; Ge, J.; Aranda, X.G.; Planz, J.V.; Eisenberg, A.J.; Chakraborty, R. Texas population substructure and estimating the rarity of Y STR haplotypes in forensic analyses. *J. Forensic Sci.*, **2009**, *54*(5), 1016-1021.

Received: January 07, 2010 Revised: March 20, 2010 Accepted: April 02, 2010

© Smalling *et al*.; Licensee *Bentham Open.*

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/-licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.