J Forensic Sci, November 2008, Vol. 53, No. 6
doi: 10.1111/j.1556-4029.2008.00880.x
Available online at: www.blackwell-synergy.com

Danielle T. Baute,[1] M.S.; Jessica A. Satkoski,[2] Ph.D.; Theresa F. Spear,[3] M.A.; David G. Smith,[2,3,4] Ph.D.; Melody R. Dayton,[1] M.S.; Venkat S. Malladi,[2] B.S.; Vivek Goyal,[2] B.S.; Alexander Kou,[2] B.S.; Jennifer L. Kinaga,[2] B.S.; and Sreetharan Kanthaswamy,[2,3,4] Ph.D.

Analysis of Forensic SNPs in the Canine mtDNA HV1 Mutational Hotspot Region*

ABSTRACT: A 60 bp sequence variation hotspot in the canine mitochondrial DNA hypervariable region 1 was evaluated for its use in forensic investigations. Nineteen haplotypes containing 18 single nucleotide polymorphisms were observed among laboratory-generated and GenBank-derived domestic dog sequences representing five regional localities in the U.S. Samples from the different localities were highly variable with the levels of intra-population variability being similar among the populations studied. AMOVA further confirmed that there was no significant genetic structuring of the populations. Assays using these haplotypes were robust, canid specific and portend a rapid method for correctly excluding individual dogs as noncontributors of forensic evidence. Species-specificity of the primers was confirmed by means of in-tube polymerase chain reaction of human and cat DNA and in-silico assessment of the genomes of several animal species. Breed-specific fragments were not detected among the common haplotypes but there is evidence that this assay may be capable of differentiating domestic dog, wolf, and coyote sequences.

KEYWORDS: forensic science, trace evidence, domestic dog, rapid screening, mitochondrial DNA, sequence variation, hypervariable region 1

The domestic dog (*Canine familaris*) mitochondrial genome is *c.* 16727 bp long (1) and most of sequence variation is in the hypervariable region 1 (HV1) and hypervariable region 2 (HV2) between nucleotide position (np) 15458 and 16727 (2).

Himmelberger et al. (3) amplified an 800 bp fragment of canine mitochondrial DNA (mtDNA) containing the entire HV1 region. This sequence can be edited to create a consensus sequence of 608 bp, which can then be compared to a reference sample. In this study, we amplified and sequenced a much shorter 145 bp amplicon containing a highly polymorphic fragment *c.* 60 bp in length between np 15595 and 15654 in the HV1 region of the canine mtDNA genome. Himmelberger et al. (3) showed that approximately a third of the 32 point mutations discovered within the 608 bp fragment were located within this 60 bp block and considered this region a variation "hotspot" for single nucleotide polymorphisms (SNPs) within the canid mtDNA (see Fig. 1).

In this study, this 60 bp region is further analyzed to determine if typing this shorter amplicon can provide sufficient discriminating information for forensic purposes. The analysis of this short region of the canine mtDNA genome could potentially provide a faster means to preliminarily screen evidence samples for exclusion purposes when candidate suspects are available. Indeed, if the sample in question does not match a suspected reference sample then further testing of the longer 608 bp fragment or an assay of the whole genome would *not* necessarily be warranted. Furthermore, this highly polymorphic 60 bp region is more suitable for investigating DNA that is highly degraded or contaminated with DNA degrading agents because amplification of regions larger than 230 bp is typically unattainable for highly degraded samples (4,5). To address this, we challenged the 60 bp assay with a degradation agent and compared the results to assays that included the analysis of the 608 bp fragment reported by Himmelberger et al. (3). We also examined aspects of the species-specificity of this test to help establish its reliability and determined basic population parameters such as haplotype diversity and distribution of the 60 bp fragment in local and global dog samples.

[1]The Jan Bashinski DNA Laboratory, California Department of Justice, Richmond, CA 94804.
[2]Department of Anthropology, University of California, Davis, CA 95616.
[3]Graduate Group in Forensic Science, University of California, Davis, CA 95616.
[4]California National Primate Research Center, University of California, Davis, CA 95616.

*Part of the research described in this article has been reported in Ms. Danielle Baute's Masters of Science (Forensic Science) thesis which will be submitted to the UC Davis Office of Graduate Studies. Ms. Baute has also presented this work at the 2007 UC Davis Graduate Group in Forensic Science Student Seminar as part of the program's requirement.

Received 18 Sept. 2007; and in revised form 15 Dec. 2007; accepted 19 Apr. 2008.

Materials and Methods

DNA Extraction

Blood samples were obtained from veterinary diagnostic laboratories in California, Colorado, Massachusetts, Ohio, and Texas. Hair root samples were collected from volunteers' dogs in California. These samples represented 30 distinct breeds recognized by the American Kennel Club (AKC). Breed information was furnished by the diagnostic laboratories that supplied the blood samples or by the dogs' owners who volunteered their pets' hair samples. DNA from hair samples was extracted using the Promega Tissue and Hair Extraction Kit (Promega, Madison, WI). DNA was extracted from 200 μL of whole blood samples using the QIAmp DNA Minikit (QIAGEN, Valencia, CA).

PCR Amplification

Primers were designed from the mtDNA reference sequence obtained from GenBank (accession no. U96639; 6) using the

© 2008 American Academy of Forensic Sciences

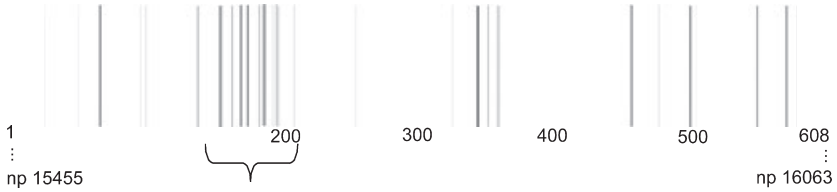

FIG. 1—*A snapshot of the aligned sixteen 608 bp haplotypes reported by Himmelberger et al. (3). Shades of grey depict the degree of nucleotide diversity. The 60 bp variation hotspot is parenthesized. The alignment was generated using the a program called* FINGERPRINT (*http://evol.mcmaster.ca/fingerprint/*).

Primer 3 website (7). The custom-designed primers were purchased (Invitrogen, Carlsbad, CA). The primer pair H15575 (5′-TTGCCCCATGCATATAAG-3′) and L15684 (5′-GTTTCTCG AGGCATGGTGAT-3′) generated a 145 bp fragment in the mtDNA HV1 control region between bases 15558 and 15703. Template DNA was amplified via polymerase chain reaction (PCR) using a reaction mixture consisting of 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 μM of forward and reverse primer, and 1 U of Platinum *Taq* polymerase (each reagent in the PCR reaction mix was purchased from Invitrogen), and 2 μL of DNA template in a total volume of 50 μL. The amplification program consisted of one cycle of 94°C for 2 min, followed by 35 cycles of denaturation at 94°C for 15 sec, primer annealing at 56°C for 30 sec, extension at 72°C for 1 min, and a final extension at 72°C for 10 min. PCR reactions were performed in an Eppendorf Gradient Mastercycler (Westbury, NY) and the presence of amplified product was verified by gel electrophoresis. Approximately 3 μL of PCR product and 3 μL sucrose were run on a 10% agarose gel and stained with ethidium bromide. The gels were run at 150 V for 30 min.

In addition to the shorter fragment analyzed here, a 608 bp fragment within the canid HV1 region previously studied in our laboratory by Himmelberger et al. (3) was subjected to the same degradation assays for comparison. The primer sequences and PCR conditions used to generate the 608 bp fragment have been described elsewhere (3).

DNA Degradation

Hair DNA extracts were subjected to degradation using a DNAse I treatment both in a Mn^{2+} buffered solution and a DNAse I solution containing no Mn^{2+} (8); both solutions were purchased from Fermentas Life Sciences (Glen Burnie, MD). While normal DNAse I produces single-stranded nicks in the double-stranded DNA (dsDNA), DNAse I in a buffer solution containing Mn^{2+} nicks the dsDNA at approximately the same site, creating blunt-ended fragments (9). Aliquots of the shorter 145 bp fragments of DNA were digested with 3 μL of DNAse I (1 U/μL) and DNAse I/Mn^{2+} (1 U/μL) respectively and at defined time intervals of 1, 2, 3, 4, 5, and 6 h. The dsDNA degradation reaction was stopped by adding 5 μL of 50 mM EDTA solution. The 608 bp HV1 fragment was initially subjected to only the DNAse I/Mn^{2+} treatment for 1–3 h to determine if it would be possible to obtain an amplified product within the expected size range. If no product was observed, no further testing with DNAse I alone was performed. For product verification, electrophoreses were performed in a 10% agarose gel stained with ethidium bromide.

Species Specificity

To test the 145 bp fragment primers for canine specificity in a mixture with human and cat DNA, one cat and one human sample were amplified with the same primer pair and PCR parameters as described above. Like human biomaterial, cat biomaterial occurs frequently in the U.S. households. Product verification was performed by electrophoresis in a 10% agarose gel and ethidium bromide staining.

The UCSC In-Silico PCR server (http://genome.ucsc.edu/cgi-bin/hgPcr?command=start) was used to run genome-wide in-silico assessment of primer specificity by simulating PCR reactions and querying the published sequences of the cat, cattle, horse, rabbit, rat, mouse, chicken, and lizard genomes for oligonucleotide sequence similarities. No matches were found when searches were conducted whether under default or broader specifications thus eliminating the potential for noncanid DNA amplification.

DNA Sequencing

The PCR product was purified by adding 0.2 μL Exo1 enzyme, 0.2 μL Exo1 buffer, the remaining PCR product, and water to a final volume of 100 μL. Samples were heated to 37°C for 90 min, then 80°C for 20 min. The samples were placed in a 96-well polystyrene plate (Millipore, Billerica, MA) and exposed to a vacuum for 10 min. The samples were rehydrated with 25 μL of water. Approximately 2 aliquots of 15 ng of product were sent to the University of California at Davis Department of Biological Sciences sequencing core facility (http://dnaseq.ucdavis.edu/) where the same PCR primer pairs were used to generate sequences from each dog sample in both forward and reverse directions.

Data Analysis

Forward and reverse sequences from the same animals were aligned to obtain consensus sequences to minimize effects from sequencing discrepancies. When a consensus strand could not be obtained, samples were resequenced and if a consensus could still not be constructed, the sample was excluded from the study. Sequences were edited and consensus strands created using BIOEDIT (10) software. Sequences were aligned to the reference dog sequence (GenBank accession no. U96639; 6) using the CLUSTALW software (11).

Eighty-three domestic dog samples were sequenced in both directions for this study. Thirteen of the samples were from California, 28 from Colorado, 10 from Ohio, seven from Massachusetts, and 25 from Texas. Thirty-six pure and mixed bred domestic dog sequences from Himmelberger et al.'s (3) California dataset (GenBank accession no. EF122413–EF122428) were also included in our study. To further increase our sampling size and regional representation, an additional 123 dog sequences that reflected over 25 breeds from Massachusetts ($n = 50$), Michigan ($n = 7$), Texas ($n = 19$), and other unspecified geographic locations were downloaded from the GenBank database. Also included in the analysis were sequences from two wolves and one coyote to assess the degree of interspecific variation within the hotspot region. The

TABLE 1—*The 126 canid sequences from GenBank.*

AY240030.1|H167	AY240063.1|H439	AY240098.1|H1687	AY240130.1|H2110
AY240031.1|H168	AY240064.1|H440	AY240099.1|H1688	AY240131.1|H2113
AY240032.1|H169	AY240065.1|H441	AY240100.1|H1689	AY240132.1|H2114
AY240033.1|H177B	AY240066.1|H442	AY240101.1|H1690	AY240133.1|H2115
AY240034.1|H177C	AY240067.1|H443	AY240102.1|H1691	AY240134.1|H2116
AY240035.1|H179B	AY240068.1|H531	AY240103.1|H1692	AY240135.1|H2118
AY240036.1|H180	AY240069.1|H532	AY240104.1|H1693	AY240136.1|H2119
AY240037.1|H195	AY240070.1|H913	AY240105.1|H1694	AY240137.1|H2120
AY240038.1|H196	AY240072.1|H915	AY240106.1|H1696A	AY240138.1|H2122
AY240039.1|H227	AY240074.1|H951	AY240107.1|H1696B	AY240139.1|H2123
AY240040.1|H229	AY240075.1|H954	AY240108.1|H1697	AY240140.1|H2124
AY240041.1|H230	AY240076.1|H956	AY240109.1|H1698	AY240141.1|H2125
AY240042.1|H231	AY240077.1|H957	AY240110.1|H1699	AY240142.1|H2126
AY240043.1|H232	AY240078.1|H958	AY240111.1|H1700	AY240143.1|H2127
AY240044.1|H233	AY240079.1|H959	AY240112.1|H1701	AY240144.1|H2132
AY240045.1|H234	AY240080.1|H967	AY240113.1|H1702	AY240145.1|H2140
AY240046.1|H287	AY240081.1|H968	AY240114.1|H1703	AY240146.1|H2141
AY240047.1|H289	AY240082.1|H970	AY240115.1|H1704	AY240147.1|H2142
AY240048.1|H290	AY240083.1|H972	AY240116.1|H1705	AY240148.1|H2143
AY240049.1|H292	AY240084.1|H975	AY240117.1|H1728	AY240149.1|H2150
AY240050.1|H293	AY240085.1|H974	AY240118.1|H1729	AY240150.1|H2157
AY240051.1|H296	AY240086.1|H980	AY240119.1|H1730	AY240151.1|H2160
AY240052.1|H313	AY240087.1|H981	AY240120.1|H1731	AY240152.1|H2165
AY240053.1|H318	AY240088.1|H982	AY240121.1|H1732	AY240153.1|H2168
AY240054.1|H324	AY240089.1|H984	AY240122.1|H1733	AY240154.1|H2178
AY240055.1|H325	AY240090.1|H985	AY240123.1|H1734	AY240156.1|H2148
AY240056.1|H326	AY240091.1|H988	AY240124.1|H1740	AY240157.1|H2151
AY240057.1|H328	AY240092.1|H993	AY240125.1|H1741	AY240073.1| C. lupus H919
AY240058.1|H333	AY240093.1|H994	AY240126.1|H1742	AY240155.1| Canis.H173
AY240059.1|H334	AY240095.1|H1684	AY240127.1|H1743	AY240094.1| Canis. H1649
AY240061.1|H426	AY240096.1|H1685	AY240128.1|H1745	
AY240062.1|H428	AY240097.1|H1686	AY240129.1|H2108	

accession numbers of all 126 GenBank sequences are provided in Table 1.

Our 119 laboratory-generated sequences were compared with the corresponding 60 bp regions in the GenBank sequences. A similarity matrix index among all sequences analyzed was generated using BIOEDIT (10). Population diversity simulations and analysis of molecular variation (AMOVA) were performed with the program ARLEQUIN version 2.001 (12). Estimates of gene diversity, pairwise nucleotide differences, and nucleotide diversity (analogs to the gene diversity of nuclear loci and expressed as π; 13), were also calculated using the ARLEQUIN software.

The exclusion capacity of the canine 60 bp sequence located in the HV1 region was estimated as $1 - \Sigma X_i^2$, where X_i is the frequency of the ith haplotype and the random match probability is $\Sigma X_i^{\ 2}$ (14). For this study, four exclusion capacity estimates were calculated: the exclusion capacity for the 119 laboratory-generated sequences including 35 sequences from Himmelberger et al. (3), estimates based on those 35 sequences alone, estimates based on the 123 GenBank dog sequences, and all 242 dog sequences, respectively.

Results

A 145 bp sequence of the canine HV1 region from 83 domestic dogs was successfully amplified. Each sequence was trimmed to interrogate the 60 bp variation hotspot which spans np 15595–15654. Combined with the variation hotspot fragments from Himmelberger et al. (3) and fragments generated from the GenBank sequences, a total of 242, 60 bp sequences representing 23 hotspot haplotypes (H1–23) were produced (Table 2).

Ten of the 23 haplotypes occurred once per state, while the rest were found in samples from 2 to 8 states; haplotype 12 was found in all eight states studied (Table 2). The sequences represented 23 domestic dog haplotypes and the frequencies at which all haplotypes were observed are also presented in Table 2. H8 and H12 were the most commonly occurring haplotypes among the 242 sequences studied at frequencies of 0.327 and 0.204, respectively. Haplotypes H2 and H10 were observed once each in the California sample from a golden retriever and a Boston terrier. Colorado haplotypes H17–H19 also occurred once each, in dogs described as a springer spaniel, a collie, and an unknown breed of dog. Haplotype H22 occurred once in Massachusetts but the breed of dog remains uncertain. Haplotypes H14 and H15 were present once each in a whippet and a dachshund mix from Texas. The location or breed of origin for haplotype H23, which is a GenBank sample, is unknown.

Fifteen haplotypes were obtained from the 123 domestic dog sequences from GenBank (Table 3). While haplotypes H1, H3–H9, and H11–H13 also occurred in the laboratory-generated dataset, four others, H20–H23, were present only in the GenBank sequences. Conversely, there were no dogs in the GenBank database that exhibited haplotypes H2, H10, and H14–H19. Therefore, between the respective datasets consisting of 242 domestic dog sequences, there were 23 domestic dog hotspot haplotypes. Among the GenBank sequences, a similar pattern of geographic distribution emerged. The more common haplotypes were found in a greater number of states. Each of the wolves and the coyote possessed different haplotypes that were absent in dogs.

Fifteen SNPs were present in 119 laboratory-generated dog sequences (Table 4). Only three additional SNPs were detected when the 123 GenBank dog sequences were included. The two wolf sequences did not contribute to the SNP tally while the coyote haplotype provided a deletion (15639^{del}), an insertion (15648^{ins}), and an additional five SNPs, making the number of canid-wide single point mutations, including the indels, a total of 25.

Sequence identity among all laboratory-generated and GenBank haplotypes ranged from a maximum of 0.983 to a minimum of

TABLE 2—*Nineteen hotspot haplotypes (1–19) obtained from the 119 laboratory-generated domestic dog sequences from this study including 10, parenthesized, from Himmelberger et al. (3) and 4 additional haplotypes (20–23) from the GenBank sequences haplotype frequencies based on all 242 sequences are also provided below.*

Haplotype	Geographic Distribution	Breeds Represented
H1 (Reference dog sequence/1) [0.087] CCTTACATAGGACATATTAACTCAATCTCATAATTCACTGATCTTTCAACAGTAATCGA	CA, CO, MA, OH, TX ma, mi	Sapsaree breed (Reference dog), *Labrador retriever, Dachshund, Labrador retriever (African), Basenji mix,* Boston terrier, Terrier mix, Beagle-Labrador mix, German shepherd, Poodle, Schnauzer, Chihuahua, French bulldog, Papillion, unknown dog breeds (old English sheepdog, Pit bull, Lhaso Apso, Poodle, Whippet, English terrier, Pug)
H2 [0.004] TCTTACATAGGACATATTCACTCAATCTCATAATTCATTGATCTGTCAGCAGTAATCAA	CA	Golden retriever
H3 (6) [0.016] CCTTACATAGGACATATTAACTCAATCTCATAGTTCACTGATCTATCAACAGTAATCAA	CA, CO, MA ma	*Labrador retriever/Golden retriever mix,* Malamute, Siberian husky (Husky)
H4 (16) [0.012] CCTTACATAGGACATATTAACTCAATCTCATAGTTCATTGATCTATCAACAGTAATCGA	CA ma	*Mongrel (Miniature schnauzer)*
H5 (12/13) [0.07] CCTTACATAGGACATATTAACTCAATCTCATAGTTCACTGATCTATCAACAGTAATCGA	CA, CO, TX ma	*Queensland heeler mix, Labrador retriever/Border collie mix,* German shepherd mix, Great Dane, Labrador retriever, Sheltie, Chow Chow, Beagle mix (Doberman, Great Dane, Great Dane, Belgium sheepdog, American Eskimo dog, Labrador retriever)
H6 (3/9/11/14) [0.135] CCTTACATAGGACATATTAACTCAATCTCATAATTCACTGATCTATCAACAGTAATCGA	CA, CO, OH, TX ma, mi	*Cairn terrier, Rottweiler mix, Labrador retriever mix, Bichon Frise, Australian shepherd mix,* Chow mix, Rottweiler, Shepherd mix, Sheltie, unknown mixed breed dogs (Springer spaniel, Chow Chow, Greyhound, Pit bull, Husky, Rottweiler, Beagle, Chihuahua, Yorkshire terrier, Anatolian shepherd, Wheaton terrier, Rottweiler, Labrador retriever)
H7 [0.008] CCTTACATAGGACATATTAACTCAATCTCATAGTTCACTGATCTGTCAACAGTAATCGA	CO ma	Irish wolfhound (Brittany spaniel)
H8 (2/4) [0.328] CCTTACATAGGACATATTAACTCAACCTCATAGTTCACTGATCTATCAACAGTAATCGA	CA, CO, MA, OH, TX ma, tx	*Labrador retriever, Airedale terrier/Golden retriever mix, Bull mastiff, Mongrel, German shepherd mix, American Pit bull Terrier,* Golden retriever, Shepherd, Maltese, Boxer, Labrador retriever, Golden retriever, Flat-coated retriever, German wirehair pointer, unknown breeds, Husky, Siberian husky, Australian shepherd, unknown breed (Golden retriever, Labrador retriever, Pit bull, Boxer, Chow Chow, Yorkshire terrier, Chesapeake Bay retriever)
H9 (15) [0.012] CCTTACATAGGACATATTAACTCAATTTCATAATTCACTGATCTATCAACAGTAATCGA	CA ma	*Mixed terrier* (American Eskimo dog, American Spitz)
H10 (5) [0.004] CCTTACATAGGACATATTAACTCAATCTCATAATTCACTGATCTTTCAACAGTAATCGG	CA	*Boston terrier*
H11 [0.008] CCTTACATAGGACATACTAACTCAATTTCATAATTCACTGATCTGTCAACAGTAACCGA	TX mi	Doberman (Doberman)
H12 (7/10) [0.205] TCTTACATAGGACATATCAACTCAATCTCATAATTCATTGATCTGTCAGCAGTAATCAA	CA, CO, MA, OH, TX ma, mi, tx	Australian Shepherd/Rottweiler mix, German Shepherd/Border Collie mix, Mini Poodle, Mongrel, Labrador retriever/Golden retriever mix, Australian shepherd, Standard Poodle mix, Golden retriever, Labrador retriever, Rottweiler, Schnauzer, Collie, Tibetan terrier mix, Poodle mix, Golden retriever, Rottweiler mix, Chihuahua, Weimereiner mix, Dachshund, unknown dog breeds (Golden retriever, Labrador retriever, Bassett hound, English bulldog, Toy poodle, Standard poodle, Kerry Blue terrier, Dachshund, Cairn terrier, Lhaso Apso, Toy poodle, German Shorthair pointer)
H13(8) [0.045] CCTTACATAGGACATACTAACTCAATCTCATAATTCACTGATCTGTCAACAGTAACCGA	CA, CO, TX ma	*Border collie,* Pit Bull, Papillion, German shepherd, unknown dog breed (Siberian husky, Border collie, Cocker spaniel, West Highland terrier, Dalmatian)

TABLE 2 *(Continued)*

Haplotype	Geographic Distribution	Breeds Represented
H14 [0.04] CCTTACATAGGAGATATTAACTCAATCTCATAATTCACTGATCTTTCAACAGTAATCGA	TX	Whippet
H15 [0.04] CCTTACATAGGACAGATTAACTCAATTTCATAATTCACTGATCTATCAACAGTAATCGA	TX	Dachshund mix
H16 [0.012] CCTTACATAGGACATATCAACTCAATCTCATAATTCATTGATCTGTCAGCAGTAATCAA	CO	Malinois, German shepherd
H17 [0.04] CCTTACATAGGACATATTAACTCAATCTCATAATTCACTGATCTATCAACAGTAACCGA	CO	Springer spaniel
H18 [0.04] CATTACATAGGACATATTAACTCAATCTCATAATTCACTGATCTATCAACAGTAATCGA	CO	Collie
H19 [0.04] TCTTACATAGGACATATCAACTCAATCTCATAATTCATTGATCTGTCAGCAGTAATCGA	CO	Unknown dog breed
H20 [0.04] CCTTACATAGGACATATTAACTCAATCTCATAGTTCACTGATCTACCAACAGTAATCGA	tx	(American Eskimo dog)
H21 [0.04] CCTTACATAGGACATATTAACTCAATCTCATAGTTCACTGATCTATCAACAGTAATCGG	tx	(Alaskan husky)
H22 [0.012] CCTTACATAGGACATACTAACTCAATCTCATAATTCACTGATCTATCAACAGTAACCGA	ma	(Shar Planinetz Anatolian shepherd)
H23 [0.04] TCTTACATAGGACATATCAACTCAATCCCATAATTCATTGATCTGTCAGCAGTAATCAA	Unknown	(Chesapeake Bay retriever)

Geographic information is based on laboratory generated and GenBank sequences: CA, California sample; CO, Colorado; MA, Massachusetts; MI, Michigan; OH, Ohio; TX, Texas. Location information in lower case (ma, mi, tx) refers to the GenBank samples.

Italicized font indicates CA breeds from Himmelberger et al. (3) and parentheses indicate samples from GenBank.

0.847 (data not shown). Haplotypes H6, H10, and H14 were most similar to the reference dog haplotype (H1) with 98.3% identity while H23 was most divergent from H1 with 88.1% identity. Among the laboratory-generated haplotypes, H2 and H12 showed the least similarity to the reference dog sequence with only 89.8% identity. H5 was most similar to H21 and H20 with 98.3% identity while H8, H11, H15, H20, H21, and H22 shared the least identity with H23 at 84.7% similarity.

Results of the AMOVA comparing the laboratory-generated and GenBank-derived haplotype data are presented in Table 5. Despite the smaller sample size, the laboratory-generated data yielded a higher degree of gene diversity (0.85) than the GenBank (0.78) data. The average pairwise difference between sequences and nucleotide diversity, however, were slightly higher in the GenBank. When only GenBank samples with known geographic origin were included, estimates of each of the parameters were much higher. Almost all of the variation (99.85%) stemmed from differences between individual samples within each of the two datasets (data not shown).

Table 6 shows a range of gene diversity estimates among dogs from different states from 0.78 (Ohio) to 0.91 (Colorado). Colorado samples also exhibited the highest number of haplotypes per sequence (12/28) despite the fact that more samples were analyzed from California (11/49), Massachusetts (11/57), and Texas (11/44). Pairwise nucleotide differences and nucleotide diversity were highest among the Texas samples (4.24 and 0.072, respectively). The lowest pairwise difference and nucleotide diversity were estimated for the Michigan and Massachusetts samples, respectively. However, the AMOVA of the different geographic samples shows that over 99.5% of the differences among all dogs used in this study were a result of matrilines within the respective locations.

When all the samples were considered in terms of haplotypic variability, a significant amount of variation (98%) was observed among individual haplotypes within each locality whereas variation among the different localities was minimal (<2%). The average gene diversity, nucleotide diversity, and pairwise differences among all dogs were 0.853 (±0.059), 0.064 (±0.04), and 3.75 (±1.91), respectively. In summary, genetic variability as measured by all the parameters was high in each population with the levels of intrapopulation variability being similar among the populations studied and relatively little variation occurred among populations from different states (Table 7).

The overall exclusion capacity based on all 242 sequences was 0.82. This suggests that in a population of 100 ''suspect'' dogs, 82 dogs could be correctly excluded as possible contributors of an unknown sample if the entire dataset was used in the analysis. Based on the 123 GenBank sequences an estimate of 0.78 was obtained while an estimate of 0.84 was obtained from the 119 laboratory-generated sequences. It is quite remarkable that an estimate of 0.82 was generated when only 36 hotspot sequences from Himmelberger et al. (3) were used. These results suggest that smaller local databases can be as useful in weighing the meaning of ''matching'' haplotypes as more regionally representative databases.

DNAse I and DNAse I/Mn^{2+} treatment of the 145 bp fragments yielded product at hours 1–4 but failed to yield product at hours 5 and 6. Similar treatment of the 608 bp fragment with the degradation agent DNAse I/Mn^{2+} yielded no product at any time point. Our species-specificity tests revealed that human and cat samples are not amplified by our primers implying that the primers were highly specific to canid DNA templates in a mixture containing human and/or cat DNA.

TABLE 3—*The 15 haplotypes derived from the GenBank domestic dog sequences. Geographic locations and breed information are provided when available.*

H1		**H8**		H2160/Labrador retriever		H177B/Border collie	MA
H195/Old English sheepdog	MA	H227/Pit bull		H2165/Labrador retriever		H177C/Border collie	MA
H227/Pit bull	MI	H292/Boxer	MA	H2168/Labrador retriever		H958/Cocker spaniel	MA
H230/Lhaso Apso	MI	H443/Chow Chow	MA	H2178/Labrador retriever		H970/West Highland terrier	MA
H287/Poodle	MA	H951/Yorkshire terrier	MA	H2151/Labrador retriever		H993/Dalmatian	MA
H318/Whippet	MA	H974/Chesapeake Bay retriever	MA				
H982/English terrier	MA	H980/Labrador retriever	MA	**H9**		**H20**	
H985/Pug	MA	H981/Boxer	MA	H957American Eskimo dog	MA	H168/American Eskimo dog	TX
		H994/Labrador retriever	MA	H975/American Spitz	MA		
H3		H1685/Golden retriever	TX			**H21**	
H984/Husky	MA	H1689/Golden retriever	TX	**H11**		H169/Alaskan husky	TX
		H1690/Golden retriever	TX	H231/Doberman	MI		
H4		H1692/Golden retriever	TX			**H22**	
H440/Miniature schnauzer	MA	H1693/Golden retriever	TX	**H12**		H179B/Shar Planinetz	MA
H972/Miniature schnauzer	MA	H1694/Golden retriever	TX	H233/Bassett hound	MI	H180/Shar Planinetz	MA
		H1696A/Golden retriever	TX	H234/English bulldog	MI	H531/Anatolian shepherd	MA
H5		H1696B/Golden retriever		H289/Toy poodle	MA		
H296/Doberman	MA	H1697/Golden retriever	TX	H293/Standard poodle	MA	**H23**	
H334/Great Dane	MA	H1698/Golden retriever	TX	H325/Kerry Blue terrier	MA	H428/Chesapeake Bay retriever	
H915/Great Dane	MA	H1701/Golden retriever	TX	H326/Dachshund	MA		
H968/Belgium sheepdog	MA	H1702/Golden retriever	TX	H442/Cairn terrier	MA		
H988/American Eskimo dog	MA	H1703/Golden retriever	TX	H913/Lhaso Apso	MA		
H2110/Labrador retriever		H1704/Golden retriever	TX	H956/Toy poodle	MA		
H2119/Labrador retriever		H1728/Golden retriever	TX	H959/German shorthair pointer	MA		
H2132/Labrador retriever		H1730/Golden retriever	TX	H1684/Golden retriever	TX		
H2148/Labrador retriever		H1731/Golden retriever	TX	H1686/Golden retriever	TX		
		H1732/Golden retriever	TX	H1687/Golden retriever	TX		
H6		H1733/Golden retriever	TX	H1688/Golden retriever	TX		
H229/Springer spaniel	MI	H1734/Golden retriever	TX	H1691/Golden retriever	TX		
H232/Chow Chow	MI	H1742/Golden retriever	TX	H1699/Golden retriever	TX		
H290/Greyhound	MA	H1743/Golden retriever	TX	H1700/Golden retriever	TX		
H313/Pit bull	MA	H2108/Labrador retriever		H1705/Golden retriever	TX		
H324/Husky	MA	H2113/Labrador retriever		H1729/Golden retriever	TX		
H333/Rottweiler	MA	H2120/Labrador retriever		H1740/Golden retriever	TX		
H426/Beagle	MA	H2123/Labrador retriever		H1741/Golden retriever	TX		
H439/Chihuahua	MA	H2124/Labrador retriever		H1745/Golden retriever	TX		
H441/Yorkshire terrier	MA	H2125/Labrador retriever		H2114/Labrador retriever			
H532/Anatolian shepherd	MA	H2126/Labrador retriever		H2115/Labrador retriever			
H954/Wheaton terrier	MA	H2127/Labrador retriever		H2116/Labrador retriever			
H967/Rottweiler	MA	H2140/Labrador retriever		H2118/Labrador retriever			
H2141/Labrador retriever		H2142/Labrador retriever		H2122/Labrador retriever			
		H2143/Labrador retriever					
H7		H2150/Labrador retriever		**H13**			
H328/Brittany spaniel	MA	H2157/Labrador retriever		H167/Siberian husky	MA		

Discussion

The frequency at which a haplotype is observed provides considerable insight into its relevance to forensic investigations. The rarer the haplotype, the greater its utility in casework because on average a more common haplotype will exclude fewer animals. Thus, for example, if an evidence sample was typed and determined to be haplotypes H12 or H8 (among the most common haplotypes), the typing result might not mean much unless the typing test resulted in an exclusion.

Himmelberger et al.'s (3) analysis of a 608 bp segment of the canine HV1 region identified 32 SNPs among 36 dogs, 12 of which were observed within the 60 bp hotspot region. In this study, only three additional SNPs were identified in this hotspot region among the 84 sequences that were generated despite a twofold increase in samples and including animals from several U.S. localities. This suggests that while the 60 bp block is a highly polymorphic region, it has a mutation rate that is relatively low. Low mutation rates are as critical for match comparisons as they are for conclusions based on cross generational population genetic data. The lower exclusion rate of only 0.83 from the 60 bp hotspot sequences generated from Himmelberger et al.'s (3) dataset as compared to previous estimates of 0.89 calculated using their 608 bp sequences is attributed to sequence length of the target DNA. Analyzing the 60 bp hotspot sequences (<10% the size of the 608 bp fragment), produced an exclusion capacity that was reduced by only 8%.

The inclusion of additional sequences from GenBank which doubled the overall sample size from 119 to 242 did not increase the exclusion rate estimates. The GenBank dataset of 123 samples yielded a much lower estimate of exclusion of 78% probably because it contained many more pure bred and related animals than the laboratory-generated dataset. However, the use of only 36 CA samples generated an exclusion capacity estimate of 82% that is identical to that based on all 242 domestic dog samples representing at least six different regional localities (California, Colorado, Massachusetts, Michigan, Ohio, and Texas) and comparable to the estimate of 84% based on the 119 laboratory-generated sequences representing five locations (California, Colorado, Massachusetts, Ohio, and Texas). Calculations of genetic diversity suggest that the observation that variation in haplotypes is significant within populations but not between populations is probably attributable to the highly variable mtDNA lineages within the respective subpopulations.

TABLE 4—*The 18 canid hotspot haplotype SNPs.*

Haplotype	15595	15596	15607	15609	15611	15612	15617	15620	15621	15622	15623	15625	15627	15628	15632	15639	15640	15643	15648	15649	15650	15651	15652	15653	15654
H1	C	C	C	T	T	T	C	T	C	T	C	T	A	T	C	T	T	A	–	A	A	T	C	G	A
H2	T	C	C	T	T	T	C	T	C	T	C	T	A	T	T	G	T	G	–	A	A	T	C	A	A
H3	C	C	C	T	T	T	C	T	C	T	C	T	G	T	C	A	T	A	–	A	A	T	C	A	A
H4	C	C	C	T	T	T	C	T	C	T	C	T	G	T	T	A	T	A	–	A	A	T	C	G	A
H5	C	C	C	T	T	T	C	T	C	T	C	T	G	T	C	A	T	A	–	A	A	T	C	G	A
H6	C	C	C	T	T	T	C	T	C	T	C	T	A	T	C	A	T	A	–	A	A	T	C	G	A
H7	C	C	C	T	T	T	C	T	C	T	C	T	G	T	C	G	T	A	–	A	A	T	C	G	A
H8	C	C	C	T	T	T	C	C	C	T	C	T	G	T	C	A	T	A	–	A	A	T	C	G	A
H9	C	C	C	T	T	T	C	T	T	T	C	T	A	T	C	A	T	A	–	A	A	T	C	G	A
H10	C	C	C	T	T	T	C	T	C	T	C	T	A	T	C	T	T	A	–	A	A	T	C	G	G
H11	C	C	C	T	C	T	C	T	T	T	C	T	A	T	C	G	T	A	–	A	A	C	C	G	A
H12	T	C	C	T	T	C	C	T	C	T	C	T	A	T	T	G	T	G	–	A	A	T	C	A	A
H13	C	C	C	T	C	T	C	T	C	T	C	T	A	T	C	G	T	A	–	A	A	C	C	G	A
H14	C	C	G	T	T	T	C	T	C	T	C	T	A	T	C	T	T	A	–	A	A	T	C	G	A
H15	C	C	C	G	T	T	C	T	T	T	C	T	A	T	C	A	T	A	–	A	A	T	C	G	A
H16	C	C	C	T	T	C	C	T	C	T	C	T	A	T	T	G	T	G	–	A	A	T	C	A	A
H17	C	C	C	T	T	T	C	T	C	T	C	T	A	T	C	A	T	A	–	A	A	C	C	G	A
H18	C	A	C	T	T	T	C	T	C	T	C	T	A	T	C	A	T	A	–	A	A	T	C	G	A
H19	T	C	C	T	T	C	C	T	C	T	C	T	A	T	T	G	T	G	–	A	A	T	C	G	A
H20	C	C	C	T	T	T	C	T	C	T	C	T	G	T	C	A	C	A	–	A	A	T	C	G	A
H21	C	C	C	T	T	T	C	T	C	T	C	T	G	T	C	A	T	A	–	A	A	T	C	G	G
H22	C	C	C	T	C	T	C	T	C	T	C	T	A	T	C	A	T	A	–	A	A	C	C	G	A
H23	T	C	C	T	T	C	C	T	C	C	C	T	A	T	T	G	T	G	–	A	A	T	C	A	A
H919/wolf	T	C	C	T	T	T	C	T	C	C	C	C	A	T	T	A	T	A	–	A	A	T	C	A	A
H173/wolf	T	C	C	T	T	C	C	T	C	T	C	C	A	T	T	A	T	A	–	A	A	T	C	A	A
H1649/cotoye	C	T	C	T	C	T	T	C	C	T	T	C	G	C	T	–	T	A	T	G	G	T	T	A	A

15 of these SNPs were found in domestic dogs.

TABLE 5—*Gene and nucleotide diversity estimates and pairwise nucleotide differences between the laboratory-generated and GenBank domestic dog sequences.*

Sample	Haplotypes/Sequences	Gene Diversity	Pairwise Difference	Nucleotide Diversity
Laboratory-generated sequence data	19/119	0.850 ± 0.017	3.59 ± 1.83	0.061 ± 0.03
GenBank sequence data	15/123 (14/74)	0.780 ± 0.027 (0.855 ± 0.02)	3.72 ± 1.89 (3.99 ± 2.02)	0.063 ± 0.036 (0.07 ± 0.04)

Estimates based on GenBank haplotypes with known geographic origin (Massachusetts, Michigan, and Texas) are in parentheses.

TABLE 6—*Gene and nucleotide diversity estimates and pairwise nucleotide differences among the different geographic localities based on the combined laboratory-generated and GenBank dog sequence data.*

Population	Haplotypes/Sequences	Gene Diversity	Pairwise Difference	Nucleotide Diversity
California	11/49	0.838 ± 0.03	3.56 ± 1.84	0.060 ± 0.035
Colorado	12/28	0.905 ± 0.03	3.66 ± 1.91	0.062 ± 0.04
Massachusetts	11/57	0.889 ± 0.02	3.46 ± 1.79	0.059 ± 0.03
Michigan	4/7	0.857 ± 0.10	4.00 ± 2.27	0.067 ± 0.04
Ohio	4/10	0.778 ± 0.09	4.022 ± 2.19	0.068 ± 0.04
Texas	11/44	0.806 ± 0.04	4.241 ± 2.14	0.072 ± 0.04
Total	22/195	0.853 ± 0.01	3.751 ± 1.91	0.064 ± 0.04

TABLE 7—*AMOVA design and results based on differences among different geographic localities based on the combined data from laboratory-generated and GenBank domestic dog sequences.*

Source of Variation	Degree of Freedom	Percentage of Variation
Among populations*	5	0.45
Within populations	189	99.55

*The six populations considered are from California, Colorado, Massachusetts, Michigan, Ohio, and Texas.

Therefore, this study shows that while larger and more geographically representative samples can increase the estimates of mtDNA genetic diversity and power of match probability based on sequence data, results also support the findings of Himmelberger et al. (3) that the local samples are adequately representative of the larger global collection. Moreover, the local dataset represents the more relevant sample from which a "suspect" dog belongs. Given that more locality-specific samples were included in this study (i.e., samples from California, Colorado, Massachusetts, Michigan, Ohio, and Texas), comparisons of variation among these samples show that samples from different localities do not show significantly different mtDNA variation. No significant population structure could be detected among the dog populations even after the GenBank samples were included in the analysis.

As observed by Himmelberger et al. (3), Angleby and Savolainen (14), and Gundry et al. (15) who used much longer mtDNA sequences, 60 bp haplotypes that occurred at higher frequencies did not correlate with breed affiliation or geographic origin of the animals. This is an important limitation of any canine mtDNA assay in forensic applications because no accurate determination of breeds or their geographic origins can be reliably established using the tests.

While breed affiliation cannot be determined even with much larger mtDNA sequences, species determination can be established using the 60 bp region, especially to distinguish between dogs and coyotes and between wolves and coyotes. Species identification is an important element in animal forensics when the suspect animal species' geographic ranges tend to overlap.

Our study suggests that this 60 bp region, along with the longer sequences described by Himmelberger et al. (3), Angleby and Savolainen (14), and Gundry et al. (15) can be used for exclusionary purposes for typing trace or degraded casework samples. However, based on its requirement of a much shorter fragment and its robustness and specificity to canid DNA and not DNA of human or feline origin, this 60 bp assay may represent a valuable tool as a rapid preliminary screening technique for excluding noncontributing dogs. When the canine DNA evidence is highly probative, the screening of the 60 bp hotspot region will require a comparison with a reference hotspot haplotype and if such a reference sample is unavailable, a full profile (see Himmelberger et al. [3]) should be developed at the time of testing.

Conclusions

Targeting the 60 bp hotspot region of the mtDNA canine genome can be an effective, robust, and rapid method to exclude individuals from a population as the source of a questioned sample. In cases where there may be many samples to analyze, the analytical time can be significantly decreased if the 60 bp hotspot region screening technique is used as a first step in the analytical process. Moreover, targeting this 60 bp hotspot will result in a higher success of amplification of degraded evidence samples than targeting larger fragments. If this region excludes an individual dog as a contributor of the questioned sample then no further testing would be needed. Our study also shows that the inclusion of more globally representative mtDNA sequences do not significantly increase the exclusion capacity estimates. Each local sample was shown to be equally highly variable and undifferentiated from each other, and the sample size is sufficiently large that it is representative of the much broader geographic sample, reiterating findings of Himmelberger et al. (3). These observations suggest that the match probability is approximately equal whether we use a local or a global reference canine database.

Acknowledgments

This research was partially funded by the National Institute of Justice (Grant No. 2004-DN-BX-K007). We would like to thank the numerous diagnostic laboratories and volunteers for donating multiple blood samples. We would like to thank Andrea Himmelberger for supplying samples from her own project. We are grateful to Dr. Robert Rice, Department of Environmental Toxicology, UC Davis for his advice on

improving DTB's thesis from which this manuscript was derived. We also wish to thank the journal's reviewers for their constructive comments.

References

1. Kim KS, Lee SE, Jeong HW, Ha JH. The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome. Mol Phylogenet Evol 1998;10:210–20.
2. Okumura N, Ishiguro N, Nakano M, Matsui A, Sahara M. Intra- and interbreed genetic variations of mitochondrial DNA major non-coding regions in Japanese native dog breeds (*Canis familiaris*). Anim Genet 1996;27:397–405.
3. 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.
4. Bender K, Farfan MJ, Schneider PM. Preparation of degraded human DNA under controlled conditions. Forensic Sci Int 2004;139:135–40.
5. Gabriel MN, Huffine EF, Ryan JH, Holland MM, Parsons TJ. Improved mtDNA sequence analysis of forensic remains using a ''mini-primer set'' amplification strategy. J Forensic Sci 2001;46:247–53.
6. Kim SK, Seong EL, Ho WJ, Ji Hong HH. The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome. Mol Phylogenet Evol 1998;102:210–20.
7. Rozen S, Skaletsky HS. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: methods in molecular biology. Totowa, NJ: Humana Press, 2000;365–86.
8. Bollum FJ. Degradation of the homopolymer complexes polydeoxyadenylate–polydeoxythymidylate, polydeoxyinosinate–polydeoxycytidylate, and polydeoxyguanylate–polydeoxycytidylate by deoxyribonuclease I. J Biol Chem 1965;240:2599–601.
9. Melgar E, Goldthwait DA. Deoxyribonucleic acid nucleases. II. The effects of metals on the mechanism of action of deoxyribonuclease I. J Biol Chem 1968;243:4409–16.
10. Hall TA. BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 1999;41:95–8.
11. Higgins D, Thompson J, Gibson T. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–80.
12. Schneider S, Roessli D. ARLEQUIN v. 2.001: a software for population genetic analysis. Geneva: Laboratory of Genetics and Biometry, University of Geneva, 2000.
13. Nei M. Molecular genetics. New York: Columbia Press, 1987.
14. Angleby H, Savolainen P. Forensic informativity of domestic dog mtDNA control region sequences. Forensic Sci Int 2005;154:99–110.
15. Gundry RL, Allard MW, Moretti TR, Honeycutt RL, Wilson MR, Monson KL, et al. Mitochondrial DNA analysis of the domestic dog: control region variation within and among breeds. J Forensic Sci 2007;52(3):562–72.

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