Peter Savolainen,[1] M.Sc.; Björn Rosén,[2] M.Sc.; Anders Holmberg,[1] M.Sc.; Thomas Leitner,[1,3] Ph.D.; Mathias Uhlén,[1] Ph.D.; and Joakim Lundeberg,[1] Ph.D.

Sequence Analysis of Domestic Dog Mitochondrial DNA for Forensic Use

REFERENCE: Savolainen P, Rosén B, Holmberg A, Leitner T, Uhlén M, Lundeberg J. Sequence analysis of domestic dog mitochondrial DNA for forensic use. J Forensic Sci 1997;42(4):593–600.

ABSTRACT: A method has been developed for the direct sequencing of hypervariable region 1 (HV1) of domestic dog (*Canis familiaris*) and wolf (*Canis lupus*) mitochondrial DNA (mtDNA) using single hairs as template. The method uses a robotic work-station and an automated sequencer to allow for robust routine analysis. A population data base was created in order to investigate the forensic and population-genetic informativeness of domestic dog HV1. Sequence variation, partitioning of dog breeds among sequence variants and phylogenetic relations between the variants were determined. Samples from 102 domestic dogs of 52 different breeds and two captive wolves were analyzed. Nineteen dog-sequence variants were found and the frequencies of the variants ranged from 1 to 21%. The calculated discrimination power of the region, i.e., the exclusion capacity, implied that nine out of ten disputed individuals can be excluded by this analysis. The sequence variants were found to cluster into four phylogenetic groups.

KEYWORDS: forensic science, DNA typing, dog, Canis familiaris, mitochondrial DNA, sequence analysis, hair

Hairs are frequently recovered from scenes of crime. Not only human hairs, but also animal hairs can be of forensic interest, and especially hairs from dogs are common evidence material. However, so far forensic examinations of animal hairs have been limited to morphological studies, which are generally recognized as one of the more difficult types of forensic examinations. A positive identification of an individual can never be attained, and only rarely can the matching of a hair with those of a disputed individual give a categorical exclusion or can a certain breed be excluded. The main reason for this is the large variation that exists not only within species and breeds but also between hairs from the same individual (1,2). The application of DNA analysis to examinations of animal hairs would therefore be a great improvement.

However, hairs found at the scene of a crime may derive from different individuals and should therefore be treated separately. Single shed hairs and hair shafts contain very small amounts (3) of mostly degraded DNA, and analysis of chromosomal DNA, e.g., microsatellite loci, is seldom successful. Mitochondrial DNA though, being present in 1000–10,000 copies per cell (4), offers a suitable target for amplification and subsequent sequence analysis of these samples (5,6). Mutations accumulate several times faster in mtDNA than in chromosomal DNA (7), and the noncoding control region contains two hypervariable regions, HV1 and HV2 (8,9), which are suitable for sequencing for identification purposes. Furthermore, mtDNA is maternally inherited and can thus be used to determine the maternal lineage of individuals (7,10,11) and the haploid mode of inheritance of the genome facilitates DNA sequence analysis. A drawback of sequence analysis compared to other forensic DNA techniques is the large amount of manual labor involved, and to render possible large scale routine sequence analysis a high degree of automation is necessary. Therefore a semi-automated system, comprising a robotic work-station and an automated DNA sequence analyzer, was developed.

The domestic dog originates from the wolf (12) and the earliest remains of domestic dog date from 10,000 to 15,000 years ago (13). Hypervariable region 1 (HV1) of the control region is the most variable part of animal mtDNA. The rate of sequence divergence of HV1 in humans has been estimated to be at most 33% per million years (14). Assuming the same rate of sequence evolution for dog mtDNA (7), the sequence divergence rate would correspond to less than one transition every 12,000 years in HV1. Thus, only a few transitions would have occurred since the time of domestication and, in the present-day dog population, almost all sequence variation within this region would originate from variations within the wolf population. If the domestication were an isolated event the sequence variation of domestic-dog HV1 would therefore be very limited. Fossil records suggest, however, a multiple domestication (13) and some variation was therefore anticipated. Because of the different geographical origins of the breeds there were also reasons to believe that some correlation between dog breeds and sequence variants would be found. Another possible source of genetic variation in dogs is hybridization between domestic dogs and wolves. However, only sporadic examples of this are known, and interbreeding between sympatric wolf and feral-dog populations was not apparent in two Italian studies (15).

This is an attempt to apply a DNA analysis method to the forensic examination of animal hairs. To render possible large scale routine analysis a semi-automated system was developed. Furthermore, this is the first major study of domestic-dog mtDNA. It shows the amount of genetic variation among and within breeds and the extent of correlation between dog breeds and sequence variants.

[1]Doctoral students, professor, and associate professor, respectively, Department of Biochemistry, Royal Institute of Technology, S-100 44 Stockholm, Sweden.
[2]Doctoral student, The National Laboratory of Forensic Science, S-581 94 Linköping, Sweden.
[3]Post doctoral student, Department of Clinical Virology, Swedish Institute for Infectious Disease Control, S-105 21 Stockholm, Sweden.
Received 19 April 1996; and in revised form 16 Oct. 1996; accepted 21 Oct. 1996.

Copyright © 1997 by ASTM International

Materials and Methods

Samples

Hairs were collected from 102 Swedish dogs of 52 different breeds at one animal hospital located in Stockholm, at one animal hospital located in Linköping, at two dog shows in the Stockholm area and at various other places in the Stockholm area and in the provinces of Jämtland and Västergötland. The collection was performed in a "semi-randomized" fashion so that the number of dogs belonging to some of the breeds was adjusted to attain a better agreement with the frequencies of breeds found in Sweden. Some breeds are still underrepresented in the survey, e.g., hunting dogs. In those cases when individuals were known to be maternally related all but one were excluded from the survey. Hairs were also collected from two wolves (one of Russian and one of Estonian descent) at the zoological gardens of Skansen, Stockholm. The samples were collected by gently plucking hairs from the animals.

Template Preparation

For shed hairs, an extraction protocol comprising proteinase K degradation, phenol extraction and a final washing step was used (16). For most of the collected hairs, which were plucked, the mtDNA could be extracted by a less laborious method described by Han et al. (17). Briefly, two centimetres of hair was placed in a PCR reaction tube with the root end down and 20 μL of methanol (pro analysi, Merck, Darmstadt, Germany) was added. The PCR tube was left overnight to allow the methanol to evaporate. PCR mixture was then added directly to the tube containing the hair, and a PCR amplification was performed.

Polymerase Chain Reaction

Two sets of oligonucleotide primers were designed from coyote and grey wolf sequences and used in a nested configuration. Primers D1 (5′-AGA GGG ACA TTA CGA GCA AGG-3′), D2 (5′-CCT AAG ACT TCA AGG AAG AAG C-3′) and D3 (*5′-TGT AAA ACG ACG GCC AGT* TTG ATG GTT TCT CGA GGC ATG G-3′) were synthesized by Operon Technologies (Alemada, Calif.) and primer D4 (5′-Biotin-CTC CAC CAT CAG CAC CCA AAG-3′) was synthesized on an automated DNA synthesis machine, Gene Assembler Plus (Pharmacia LKB, Uppsala, Sweden), according to the manufacturer's directions. The positions of the 3′-ends of primers D3 and D4 correspond to positions H16,403 and L15,995 respectively in the human mtDNA genome (18). Primer D3 contains a handle consisting of the M13-21 universal sequencing primer sequence (indicated by italics) to enable the use of dye-labeled M13-21 sequencing primers. Primer D4 has a biotin molecule incorporated to render possible immobilization of the amplification product onto solid support. The outer amplification was performed with primers D1 and D2. One microlitre of product from the outer amplification was used as a template in the inner amplification where primers D3 and D4 generated a 319 bp fragment. The PCR mixture of both amplifications consisted of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Tween 20, 0.2 mM of each dNTP, 0.1 μM of each primer and 1 unit of Ampli*Taq* DNA polymerase (Perkin Elmer-Cetus, Emeryville, Calif.) in a total volume of 50 μL. The PCR reactions were performed in a Perkin Elmer 9600 thermocycler (Perkin Elmer-Cetus). The outer amplification program consisted of a predenaturation step (94°C, 2 min) followed by 25 cycles of denaturation (94°C, 15 s), primer annealing (65°C, 30 s) and extension (72°C, 1 min). The inner amplification program consisted of 25 cycles of denaturation (94°C, 15 s), primer annealing (69°C, 30 s) and extension (72°C, 1 min) followed by a final extension step (72°C, 10 min). Four microlitres of the amplification product was analyzed by agarose gel electrophoresis and ethidium bromide staining for the presence of the desired fragment.

Automated Solid Phase Sequencing and Sequence Analysis

The immobilization of the PCR products onto solid support and the solid-phase sequencing reactions were performed using an ABI Catalyst robotic workstation (Applied Biosystems, Inc., Foster City, Calif.) controlled by a program developed at the Department of Biochemistry, Royal Institute of Technology (19). Briefly, 200 μg of streptavidin-coated paramagnetic beads, Dynabeads™ M-280 (Dynal A.S., Oslo, Norway), were washed with 50 μL binding solution (10 mM Tris-HCl, pH 7.5; 1 mM EDTA; 2 M NaCl) and suspended in 40 μL binding solution. Immobilization of the amplification product was achieved by mixing 40 μL of the product containing incorporated biotin primer D4 with the beads and incubating for 15 min at 20°C. Single-stranded DNA was obtained by denaturing the immobilized double-stranded DNA with 50 μL of 0.1 M NaOH for 5 min at 20°C. The NaOH supernatant, containing the eluted single strand, was removed. The beads, containing the immobilized single-stranded DNA, were washed once with 50 μL binding solution and once with 50 μL 1 × TE buffer (10 mM Tris-HCl, pH 7.5; 1 mM EDTA). The beads were finally resuspended in 18 μL of annealing buffer (56 mM Tris-HCl, pH 7.5; 20 mM $MgCl_2$) and split into four separate aliquots; 3.3 μL, 1.7 μL, 6.4 μL, and 6.4 μL for the A, C, G, and T reactions respectively. To the A aliquot 0.4 pmol of fluorescently dyed −21 M13 JOE primer (Applied Biosystems, Inc.) was added, to the C aliquot 0.3 pmol of FAM primer, to the G aliquot 0.8 pmol of TAMRA primer, and to the T aliquot 0.8 pmol of ROX primer. Annealing of the primers was achieved by heating at 65°C for 5 min followed by cooling to 25°C over 8 min after which the reactions were cooled to 4°C. In parallel, four nucleotide mixtures (1 mM dATP, 1 mM dCTP, 1mM dTTP, 1mM c7dGTP, 5 μM specific ddNTP, 0.05 M NaCl, 0.04 M Tris-HCl pH 7.5) were mixed with extension buffer (300 mM citric acid pH 7.0, 318 mM DTT, 40 mM $MnCl_2$) in a ratio of 3:1 and cooled to 4°C. These mixtures, 2, 1.5, 4, and 4 μL respectively, were added to the A, C, G, and T reactions. Also, to the A and C reactions 1 unit of T7 DNA polymerase was added, and to the G and T reactions 2 units were added. The extension reaction was performed at 37°C for 5 min and interrupted by cooling to 4°C and adding 40 μL 10 × TE buffer. The A, C, G, and T reactions were pooled and the supernatant was discarded. The beads, containing single-stranded template DNA with complementary extension products, were washed once with 50 μL 1 × TE and finally suspended in 6 μL of formamide. The reactions were taken from the robotic workstation, heated at 95°C for 2 min and subsequently directly put on ice. The reactions were analyzed on an Applied Biosystems 373A DNA Sequencer according to the manufacturer's directions. The sequences were studied for the presence of polymorphic nucleotide positions using the SeqEd software (Applied Biosystems, Inc.).

Phylogenetic Analysis

Phylogenetic analyses were performed using the PHYLIP package (20) and PAUP (21). In PHYLIP, programs DNAML (maximum likelihood) and DNADIST together with FITCH (Jukes-Cantor + Fitch-Margoliash) were used to create trees. Bootstrap

replicates were made with SEQBOOT and analyzed with DNAPARS as well as DNADIST. PAUP was used to calculate g1 statistics (22) and unweighted parsimony.

Results

The developed method is based on solid-phase sequencing of PCR products using paramagnetic beads as solid support (23). A semi-automated approach was used to enable robust routine analyses. To obtain DNA from shed hairs a standard phenol extraction method (16) was used, whereas for plucked hairs a less laborious method (17) could be used. The PCR was performed in two steps using nested primers to increase sensitivity and to increase sequence quality by reducing the amount of interfering artefact fragments. One of the inner primers was biotinylated to render possible immobilization of the product onto paramagnetic beads. By denaturing the immobilized double-stranded DNA using 0.1 M NaOH and washing away the supernatant, an extremely pure sequencing template is obtained. The use of pipettable paramagnetic beads as solid support made it possible to perform the sequencing reactions by means of a robotic workstation, thus reducing the amount of manual labor considerably. Electrophoresis and sequence analysis were performed on an automated DNA sequencer.

Genetic Variation

A 257-bp segment of the control region, containing Hypervariable region 1 (HV1), was sequenced in 102 domestic dogs of 52 different breeds (Fig. 1). A total of 19 sequence variants were found, among which 23 nucleotide positions were polymorphic (Fig. 2). Twenty-one positions showed only transitions and one (position No. 209) showed both transitions and transversions. One of the sequence variants (No. 10) had a 1-bp insertion whereas sequence variant No. 13 contained a 67-bp tandem repeat. The frequencies of the sequence variants varied between 20.6% and less than 1% (one individual per sequence variant) (Fig. 3). The three most common variants (Nos. 3, 4, and 5) were present in 51% of the individuals and the six most common variants were found in more than 75% of the individuals. To obtain a measure of the discrimination power of the region, the exclusion capacity, defined as one minus the sum of the squares of the variant frequencies, was calculated. It was found to be 0.88, which implies that on average 88 disputed individuals out of 100 can be excluded by this analysis. This figure can be compared to the results of a study by Piercy et al. (24) in which HV1 of 100 British white Caucasian humans had an exclusion capacity of 0.97. The mean pairwise

```
        10         20         30         40         50

CTGAGATTCT TCTTAAACTA TTCCCTGACA CCCCTACATT CATATATTGA     50
ATTACCCCTA CTGTGCTATG TCAGTATCTC CAGGTAAACC CTTCTCCCCT    100
CCCCTATGTA CGTCGTGCAT TAATGGTTTG CCCCATGCAT ATAAGCATGT    150
ACATAATATT ATATCCTTAC ATAGGACATA TTAACTCAAT CTCATAGTTC    200
ACTGATCTAT CAACAGTAAT CGAATGCATA TCACTTAGTC CAATAAGGGC    250
TTAATCA                                                   257
```

FIG. 1—*The DNA sequence of dog mitochondrial DNA, hypervariable region 1 of the control region. The sequence is derived from one of the individuals and is used as a reference sequence in the list of the polymorphic nucleotide positions in Fig. 2. Polymorphic nucleotide positions are marked by underlined letters. The first nucleotide corresponds to position 15,996 of the human mtDNA genome (18) and the last one corresponds to position 16,402.*

sequence distance among the dogs was 1.6%. Two wolves were also analyzed. One wolf (W1) had a sequence variant (No. 6) also found among the dogs whereas the other (W2) had a unique sequence variant which was divergent in a position (nucleotide pos. 162) not found to be variable among the dogs investigated.

Breed-Sequence Correlations

No general correlation between dog breed and sequence variant was found (Figs. 3 and 4). However, for some breeds one sequence variant was over represented. Among the various Retriever breeds, 9 out of 15 (60%) were of the same variant (No. 4) and among the Labrador Retrievers, 5 out of 6 (83%) were of variant No. 4. In the total population there were 15% Retrievers while 43% of the individuals of variant No. 4 were Retrievers. All three Afghan Hounds were of the same variant (No. 6). Of the German Shepherd Dogs, 4 out of 7 (57%) were of variant No. 5. Variant No. 8 was found only in 4 dogs belonging to two closely related Northern Scandinavian breeds, Jämthund and Norwegian Elkhound, Grey (One of the dogs is a crossbreed but has a maternal grandmother which is a Norwegian Elkhound). One Chesapeake Bay Retriever was included in the study, and it was found to have a unique sequence (variant No. 13) containing a 67-bp insertion which is a tandem repeat of nucleotides 101–167. The mother and three siblings of the dog and one unrelated dog of the same breed were therefore analyzed and were all found to have the same insertion. Except for this insertion the sequence was identical with sequence variant No. 4, which was present in 60% of the Retrievers, and it is probable that sequence variant No. 13 is a modification of variant No. 4. Wolf W1 (Russian descent) had a sequence identical with dog-sequence variant No. 6 while wolf W2 (Estonian descent) had a unique sequence variant.

Phylogenetic Analysis

Phylogenetic tree analysis was used to compare 18 of the 19 dog-sequence variants and the two wolf variants with each other; dog variant No. 13 containing the 67-bp insertion was excluded. The analyzed segment consisted of 253 nucleotide positions among which only 22 showed point mutations, making solid phylogenetic inference difficult. However, the phylogenetic signal calculated with g1 statistics had a value of −0.74, giving confidence limits greater than 99% that the data are non-random (22). Figure 5 shows a distance matrix based tree (Jukes-Cantor distance based matrix and Fitch-Margoliash calculated tree). Trees calculated using other algorithms (maximum likelihood, unweighted parsimony and neighbor joining) gave similar results (data not shown). In the trees, five groups of sequence variants could be distinguished. Of these, the largest group consisted of 76.5% of the analyzed dogs (variants 1–5, 9, 11, 12, 14–18), a second group consisted of dog variants 7 and 19 and a third one of dog variants 6 and 10 together with wolf-sequence variant W1 which was identical with dog variant No. 6. The remaining two sequences (dog variant 8 and wolf variant W2) formed branches of their own. Bootstrap analysis strongly supported only the second (99–100%) and the third (66–78%) group clusters, while the other groups gave values in the range of 30% to 50%. Bootstrap values >70% have been suggested to represent confidence limits of approximately 95% in most situations (25). The grouping of sequence variants into lineage clusters is further highlighted by the fact that the average pairwise sequence differences between groups were 2.4–4.0% while the amount of sequence difference within groups was 0.4–1.2%. The correlation between dog breed and tree group was

Variant	Polymorphic nucleotide positions																						
	5	22	30.1	53	78	96	103	123	155	(162)	165	181	182	190	195	197	202	204	206	209	213	220	222	223
1	G	T	-	T	C	C	C	A	A	T	C	T	T	T	T	G	C	G	T	A	A	T	G	A
2	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
4	.	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
5	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	T	.	.	.	.
6	A	.	.	C	.	T	.	.	.	.	T	.	C	.	.	A	T	.	.	G	G	.	A	.
7	A	.	.	C	T	T	.	.	.	.	.	C	.	.	.	A	.	.	.	G	.	C	.	.
8	A	.	.	C	.	.	.	.	.	.	.	.	.	.	C	A	T	A	C	G	G	.	.	.
9	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.
10	A	.	C	C	.	T	.	.	.	.	.	.	C	.	.	A	T	.	.	G	G	.	A	.
11	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
12	.	.	.	C	.	.	.	.	G	.	.	.	.	.	.	A	.	.	.	T	.	.	.	.
13*	.	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
14	.	.	.	C	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
15	.	C	.	C	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	T	.	.	.	.
16	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	A	.	.	.	T	.	.	.	.
17	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	T	.	.	.	G
18	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
19	A	.	.	C	T	T	T	.	.	.	.	C	.	.	.	A	.	.	.	G	.	C	.	.
W2	A	.	.	C	.	T	.	.	.	C	T	.	.	.	.	A	.	.	.	.	.	.	A	.
	5	22	30.1	53	78	96	103	123	155	(162)	165	181	182	190	195	197	202	204	206	209	213	220	222	223

FIG. 2—*List of the 23 polymorphic nucleotide positions of the 19 sequence variants found among the 102 examined dogs. The nucleotides are numbered according to the sequence of one individual used as a reference, and only differences from this sequence are shown for the various sequence variants. The numbering begins at the 3'-end of primer D4. One additional nucleotide, found in sequence variant No. 10 but not in the reference sequence, is numbered with a decimal number. The corresponding deletion in the reference sequence is indicated by a "-." Sequence variant No. 13, marked "*," contained an insertion of a 67-bp tandem repeat (nucleotides 101–167). A 24th polymorphic nucleotide position (No. 162), indicated by figures in parenthesis, was found in wolf-sequence variant W2.*

low (Fig. 4). Forty-four out of the 52 breeds were represented in the largest tree group, and in tree groups II and III a wide range of breeds were found. Among the 23 breeds represented by more than one individual, eight were found in two different tree groups. The only strong breed-tree group correlation was found in tree group IV which consisted only of Scandinavian breeds, Jämthund, and Norwegian Elkhound.

Discussion

We have described an approach to utilizing DNA analysis for examinations of dog hairs, which are a common type of forensic evidence. Up to now analyses of this type of material have been restricted to morphological examinations, which generally can give only limited information. The results obtained from the population data base show that there is sufficient sequence variation within HV1 of the dog mtDNA control region to make sequence analysis of this segment a useful forensic tool. An exclusion capacity of 0.88 implies that 9 out of 10 individuals can be excluded from an investigation. Future sequence analysis of the second hypervariable region of the D-loop, HV2, will probably enhance the discrimination power of the method.

DNA sequence analysis is very labor-intensive compared to other forensic DNA-analysis methods e.g., microsatellite analysis and DNA fingerprinting. Therefore, to create a method suitable for routine laboratories where high throughput is necessary, a major concern was to minimize the amount of manual labor. The use of a robotic work station to perform the solid-phase sequencing reactions drastically reduces the amount of manual work and it also reduces the risk for laboratory errors e.g., sample mix-up. Together with automated sequencing it creates a robust semi-automated system suitable for routine forensic investigations.

Among the 22 nucleotide positions where point mutations was found, 21 showed only transitions whereas in one position (No. 209) both transitions and transversions occurred. This is in accordance with a general pattern of transition bias in mtDNA. Transitions are found to outnumber transversions by a ratio of approximately 10:1 in a variety of species (26). In the phylogenetic analysis, sequence variants containing this transversion clustered together. Furthermore, nucleotide position No. 209 shows a high degree of variability and can therefore be used as a hotspot. By single nucleotide analysis of this position (e.g., mini-sequencing (27)) an exclusion capacity of 0.58 can be obtained according to our data, that is, more than 50% of disputed individuals can be excluded.

In this study we analyzed dogs from 52 of the most common breeds in Sweden, thereby trying to obtain values of the sequence-variant frequencies of the total dog population. However, as is indicated by this study, the frequencies of sequence variants and the exclusion capacities may differ significantly between different breeds. Therefore, in forensic casework, a minor reference data base of sequence variants from dogs of the same breed as the disputed individual should be created to obtain values of the sequence-variant frequencies of that particular breed. Furthermore, a maxim of modern dog breeding is to maintain the purity of the breed and therefore crossbreeding is rare. Thus, the foreign breeds that have been introduced into Sweden have been kept pure, which implies that similar results would be expected for other populations.

The correlation between dog breed and sequence variant was found to be low in the material. Though some sequence variants were clearly overrepresented in certain breeds, e.g., Labrador Retriever (variant No. 4) and Afghan Hound (variant No. 6), and one variant (No. 8) was found only in two Scandinavian breeds, the partitioning of dog breeds among sequence variants and phylogenetic groups was limited. This lack of correlation between sequence variant and dog breed may be surprising considering the different geographical origins of the breeds and the extent of

Variant	Breed	Number	Freq.%
1	Border Collie	2	9.8
	Irish Setter	2	
	Chow-Chow	1	
	Collie	1	
	Leonberger	1	
	Norwegian Buhund	1	
	Samoyede	1	
	Crossbreed	1	
2	Chow-Chow	1	2.9
	Tibetan Terrier	1	
	Chinese crested	1	
3	Rottweiler	2	15.7
	English Setter	2	
	Norw. Elkhound, Grey	2	
	Border Terrier	2	
	Papillon	1	
	Whippet	1	
	Springer Spaniel	1	
	Poodle	1	
	Icelandic Sheepdog	1	
	Fox Terrier	1	
	Japanese Spitz	1	
	Siberian Huskey	1	
4	Labrador Retriever	5	20.6
	Golden Retriever	2	
	Flat-coated Retriever	2	
	Norfolk Terrier	2	
	Papillon	1	
	Newfoundland	1	
	Alaskan Huskey	1	
	Giant Schnauzer	1	
	Boxer	1	
	Leonberger	1	
	Kuvasz	1	
	Schipperke	1	
	German Shepherd Dog	1	
	Crossbreed	1	
5	German Shepherd Dog	4	14.7
	English Setter	2	
	Dachshund, Wirehaird	2	
	Samoyede	1	
	Border Collie	1	
	Wachtelhund	1	
	Old English Sheepdog	1	
	Keeshond	1	
	Hamiltonstövare	1	
	Crossbreed	1	
6	Afghan Hound	3	8.8
	Golden Retriever	2	
	Groenendael	1	
	Irish Water Spaniel	1	
	Otter Hound	1	
	Tibetan Spaniel	1	
(W1	Russian wolf	1)	
7	Giant Schnauzer	2	6.9
	Alaskan Huskey	1	
	German Shepherd Dog	1	
	Airedale Terrier	1	
	Jämthund	1	
	West Highl. white Terrier	1	
8	Jämthund	2	3.9
	Norw. Elkhound, Grey	1	
	Crossbreed (Elkhound grandmother)	1	
9	Irish Setter	1	3.9
	S:t Bernhard	1	
	Greyhound	1	
	Tibetan Terrier	1	
10	Flat-coated Retriever	1	2.0
	Dachshund, Wirehaired	1	
11	Irish Wolfhound	2	2.9
	Pyrenean Mastiff	1	
12	Labrador Retriever	1	1.0
13	Chesapeake Bay Retriever	1	1.0
14	Fox Terrier, Wire	1	1.0
15	Golden Retriever	1	1.0
16	Norwegian Lundehund	1	1.0
17	Cav. King Charles Spaniel	1	1.0
18	Siberian Huskey	1	1.0
19	German Shepherd Dog	1	1.0
(W2	Estonian wolf	1)	

FIG. 3—*List of the 102 dogs examined and their distribution among the sequence variants. The breeds, the number of individuals of each breed and the sequence-variant frequencies are shown. The two wolves are listed but not included in the frequency calculations.*

breeding that they have been subject to. Founder effect and fixation of mtDNA variants through drift could have been expected for the different breeds. However, there has been a considerable amount of crossbreeding involved in the early development of many breeds, which may have had great influence on the distribution of sequence variants among the breeds. The fact that sequence variant No. 8 was only found in Scandinavian breeds indicates that analyses of other geographically isolated breeds may reveal more sequence variants that are strongly correlated with certain breeds. One of the wolves in the study had a sequence variant identical with one of the dog variants. This sequence variant may either be a "true" wolf-sequence variant or a dog variant incorporated into the wolf

Tree group:	I														II		III		IV
Sequence variant:	1	2	4	9	11	18	13	3	14	5	12	15	16	17	7	19	6	10	8
Afghan Hound																	3		
Boxer			1																
Chinese Crested		1																	
Chow-Chow	1	1																	
Collie	1																		
Collie, Border	2									1									
Dachshund, Wirehaird										2								1	
German Shepherd Dog			1							4					1	1			
Giant Schnauzer			1												2				
Greyhound				1															
Groenendael																	1		
Hamiltonstövare										1									
Huskey, Alaskan			1												1				
Huskey, Siberian						1		1											
Icelandic Sheepdog								1											
Irish Wolfhound					2														
Japanese Spitz								1											
Jämthund															1				2
Keeshond										1									
Kuvasz			1																
Leonberger	1		1																
Newfoundland			1																
Norwegian Buhund	1																		
Norwegian Elkhound, Grey								2											1
Crossbreed (Elkhound Grandmother)																			1
Norwegian Lundehund													1						
Old English Sheepdog										1									
Otter Hound																	1		
Papillion			1					1											
Poodle								1											
Pyrenean Mastiff					1														
Retriever, Chesapeake Bay							1												
Retriever, Flat-coated			2															1	
Retriever, Golden			2									1					2		
Retriever, Labrador			5								1								
Rottweiler								2											
S:t Bernard				1															
Samoyede	1									1									
Schipperke			1																
Setter, English								2		2									
Setter, Irish	2			1															
Spaniel, Cav. King Charles														1					
Spaniel, Irish Water																	1		
Spaniel, Springer								1											
Terrier, Airedale															1				
Terrier, Border								2											
Terrier, Fox								1	1										
Terrier, Norfolk			2																
Terrier, West Highland White															1				
Tibetan Spaniel																	1		
Tibetan Terrier		1		1															
Wachtelhund										1									
Whippet								1											
Crossbreed										1									
Crossbreed	1																		
Crossbreed			1																

FIG. 4—*Distribution of individuals of the different dog breeds with respect to phylogenetic tree group and sequence variant. The four main tree groups are indicated by Roman numerals whereas the sequence variants are indicated by Arabic numerals. The sequence variants of tree group I are arranged according to their distribution among the subclusters of the group.*

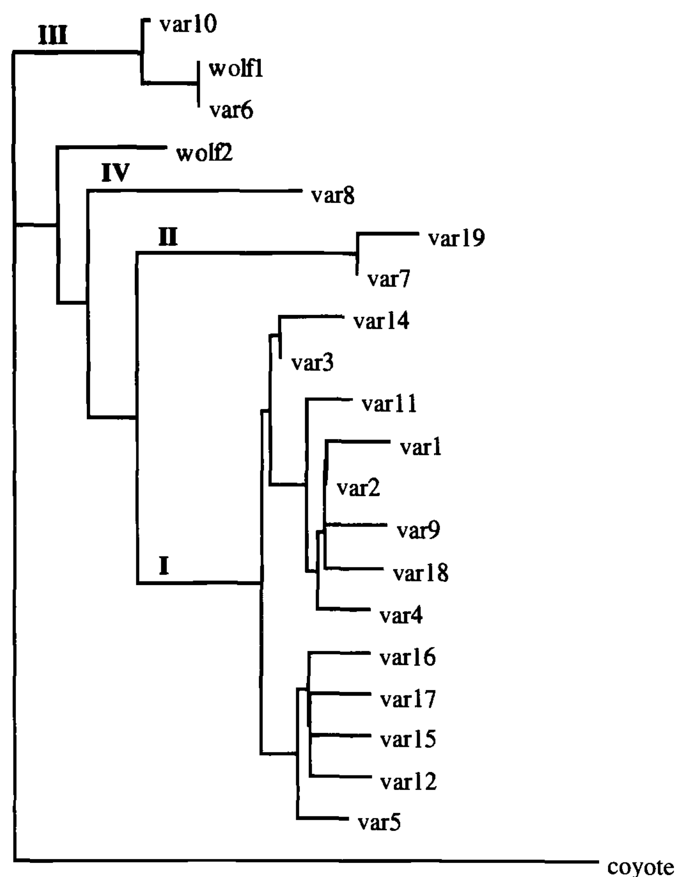

FIG. 5—*Distance matrix based tree showing phylogenetic relations between 18 of the 19 dog-sequence variants and the two wolf-sequence variants found in this work. The four dog groups are indicated by Roman numerals. The tree is rooted by a coyote sequence. The distance of the coyote sequence is reduced by 50%.*

lineage through hybridization between a wolf and a dog. In studies of the Swedish wild wolf population (data not shown) three sequence variants were found, all of which were divergent from the dog sequences, and it may now be possible to distinguish between wolf and dog in cases of suspected poaching.

Apart from the forensic field of application, the data obtained from the population data base featured interesting population genetic qualities. Phylogenetic tree analysis was used to compare the sequence variants with each other. The two hypervariable regions of the control region (HV1 and HV2) show an unusual pattern of evolution, which involves extensive formation of insertions and deletions and generation of short repeats (28). An example of tandem-repeat formation was found in sequence variant No. 13 in this study. However, we performed a phylogenetic grouping of sequence differences analyzing only point mutations. The dog variants were found to cluster into four groups with little internal variation. Within the groups the largest distance to a common ancestor was maximally 0.8%, while the average pairwise sequence difference between groups was 2.4–4.0%. It can therefore be hypothesized that the sequence difference between groups has evolved before the domestication of the wolf, and that the sequence difference within groups has evolved after the domestication took place. This indicates that the domestic dog originates from at least four female wolf lineages. Further comparative population genetic analyses of dogs and wolves might give more knowledge about the domestication of the wolf and the subsequent evolution of the domestic dog.

In conclusion, a method utilizing DNA sequence analysis for forensic examination of dog hairs was developed. To make the system attractive for routine forensic investigations, a robotic workstation, performing the DNA sequencing reactions, was integrated with a standard automated sequencer thus considerably reducing the amount of manual work and the risk for laboratory errors. The results from the population data base show that there is sufficient sequence variation within HV1 of the dog mtDNA control region to make sequence analysis of this segment a useful forensic tool.

Acknowledgments

We would like to thank the many collaborators, especially Djursjukhuset Albano, Linköpings djursjukhus, Svenska Kennelklubben, and Skansen, who provided samples. We are also grateful to Dr. Robert K. Wayne for providing coyote and grey wolf sequences, Dr. Svante Pääbo for valuable discussions and Dr. Staffan Bergh for computer assistance. This work was supported by grants from the Swedish Research Council for Engineering Sciences (TFR).

References

1. Moore JE. A key for the identification of animal hairs. J Forensic Sci Soc 1988;28:335–9.
2. Seta S, Sato H, Migake B. Forensic hair investigation. In: Maehly A, Williams RL, editors. Forensic Science Progress 2. Berlin: Springer Verlag, 1988.
3. Hukkelhoven MWAC, Vromans E, Markslag AMG, Vermorken AJM. A simple fluorimetric microassay for DNA in hair follicles or fractions of hair follicles. Anticancer Res 1981;1:341–3.
4. Bogenhagen D, Clayton DA. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. J Biol Chem 1974;249:7991–5.
5. Higuchi R, von Beroldingen CH, Sensabaugh GF, Erlich HA. DNA typing from single hairs. Nature 1988;332:543–6.
6. Wilson MR, Polanskey D, Butler J, DiZinno JA, Replogle J, Budowle B. Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. BioTechniques 1995; 18:662–9.
7. Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM, et al. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J the Linnean Soc 1985;26:375–400.
8. Greenberg BD, Newbold JE, Sugino A. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Genetics 1983;21:33–49.
9. Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC. Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 1989;86:9350–4.
10. Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci USA 1980;77:6715–9.
11. Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, et al. Identification of the remains of the Romanov family by DNA analysis. Nat Genet 1994;6:130–5.
12. Wayne RK. Molecular evolution of the dog family. Trends Genet 1993;9:218–24.
13. Olsen SJ. Origins of the domestic dog, the fossil record. Arizona: The University of Arizona press, 1985.
14. Ward RH, Frazier BL, Dew-Jager K, Pääbo S. Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci USA 1991;88:8720–4.
15. Randi E, Lucchini V, Francisci F. Allozyme variability in the Italian wolf (Canis lupus) population. Heredity 1993;71:516–22.
16. Gill P, Jeffreys AJ, Werrett DJ. Forensic application of DNA 'fingerprints'. Nature 1985;318:577–9.
17. Han C-Y, Lin BK, Lin HJ. Methanol for preparing hair bulbs for PCR. Nucleic Acids Res 1992;20:6419–20.
18. Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature 1981;290:457–65.
19. Holmberg A, Fry C, Uhlén M. Automatic preparation of DNA

templates for sequencing on the ABI Catalyst robotic workstation. In: Adams MD, Fields C, Venter JC, editors. Automated DNA sequencing analysis. London: Academic Press Limited, 1994; 139–45.
20. Felsenstein J. PHYLIP: Phylogeny inference package. Version 3.52c. Seattle, Washington: University of Washington, 1993.
21. Swofford DL. PAUP: Phylogenetic analysis using parsimony. Version 3.1.1. Champaign, Illinois: Illinois Natural History Survey, 1991.
22. Hillis DM, Huelsenbeck JP. Signal, noise and reliability in molecular phylogenetic analysis. J Hered 1992;83:189–95.
23. Hultman T, Ståhl S, Hornes E, Uhlén M. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. Nucleic Acids Res 1989;17:4937–46.
24. Piercy R, Sullivan KM, Benson N, Gill P. The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Leg Med 1993;106:85–90.
25. Hillis DM, Bull JJ. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 1993;42:182–92.
26. Simon C. Molecular systematics at the species boundary: Exploiting conserved and variable regions of the mitochondrial genome of animals via direct sequencing from amplified DNA. In: Hewitt GM, et al. editors. NATO ASI Series, Vol. H 57 Molecular Techniques in Taxonomy. Berlin, Heidelberg: Springer Verlag, 1991;33–71.
27. Syvänen AC, Aalto-Setälä K, Harju L, Kontula K, Söderlund H. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. Genomics 1990;8:684–92.
28. Saccone C, Pesole G, Sbisá E. The main regulatory region of mammalian mitochondrial DNA: Structure-function model and evolutionary pattern. J Mol Evol 1991;33:83–91.

Additional information and reprint requests:
Professor Mathias Uhlén
Department of Biochemistry, Royal Institute of Technology
Teknikringen 34
S-100 44 Stockholm, Sweden
E-mail address: mathias@biochem.kth.se