Peter Savolainen,[1] *Ph.D.; Lars Arvestad,*[2] *Ph.D.; and Joakim Lundeberg,*[1] *Ph.D.*

A Novel Method for Forensic DNA Investigations: Repeat-Type Sequence Analysis of Tandemly Repeated mtDNA in Domestic Dogs*

REFERENCE: Savolainen P, Arvestad L, Lundeberg J. A novel method for forensic DNA investigations: repeat-type sequence analysis of tandemly repeated mtDNA in domestic dogs. J Forensic Sci 2000;45(5):990–999.

ABSTRACT: A highly variable and heteroplasmic tandem repeat region situated in the mitochondrial mt DNA control region (CR) in domestic dogs and wolves was studied to evaluate its suitability as a forensic genetic marker for analysis of single hairs. The tandem repeat array is composed of three 10-bp repeat types that are distributed so that a secondary DNA sequence is formed. Thus, the region presents two levels of variation: variation in the number of repeats and variation in the secondary DNA sequence of repeat types. Two analysis methods were therefore tested; fragment length analysis and analysis of the sequence of repeat types. Fragment analysis produced unique profiles that could be used to discriminate between blood samples from maternally closely related individuals. However, different hairs from one individual did not have the same fragment profile, and the method is, therefore, not suitable for analysis of single hairs. In contrast, analysis of the repeat type sequences (array types) is highly informative. When different hairs from one individual were studied, identical array types were found. The repeat-type sequence variation was studied among individuals having identical nonrepetitive CR mtDNA sequence variants. Seven, six, and two individuals, representing three different sequence variants, respectively, were analyzed. All these individuals had different array types, which implies a very high genetic variation between individuals in this region. The analysis method considerably improves the exclusion capacity of mtDNA analysis of domestic dogs compared with sequence analysis of non-repetitive DNA.

KEYWORDS: forensic science, DNA typing, dog, wolf, *Canis familiaris,* mitochondrial DNA, tandem repeat, sequence analysis, hair

DNA analysis has become well established for forensic identification of most types of human biological material since it was first used in a forensic case in 1986 (1). In most cases, when relatively large amounts of intact DNA is present, repetitive elements situated in the nuclear DNA (microsatellites and minisatellites) are analyzed by short tandem repeat (STR) analysis (2–4) and DNA fingerprinting (1,5,6), respectively. For materials containing very small amounts of DNA or severely degraded DNA, such as shed hairs, mtDNA analysis is normally the choice. The reason for this is that the presence of more than 1000 copies of mtDNA per cell (7,8), compared with two copies of nuclear DNA loci, enhances the chances of extracting and successfully analyzing the DNA molecules. mtDNA analysis is performed by determining the DNA sequence either by conventional DNA sequence analysis (9–11) or by mini-sequencing (12). The drawback of mtDNA analysis is that the exclusion capacity is limited compared with that of nuclear DNA analysis, because of the limited information content in the small mtDNA molecule.

Biological material from animals is also frequently found at crime scenes. The most frequent type of material is hairs from furred animals, especially pets such as domestic dogs and cats. However, biological materials from animals have so far been used for the discrimination between individuals in a very limited number of forensic cases (13,14). A previous study has shown that the amount of mtDNA sequence variation is more limited in domestic dogs than in humans (15). The exclusion capacity for the most variable part of the domestic dog mtDNA control region (CR) was 0.87 in a Swedish dog population, which can be compared to a value of 0.97 found in the same region in a human population of English white Caucasians (16). The lower variation among dogs can be explained by the limited number of animals that was involved in the domestication of the wolf (17), and the limited amount of time that has passed since that event. The relatively low exclusion capacity of mtDNA sequence analysis in domestic dogs constitutes a disadvantage in its application to forensic analyses.

In the mtDNA of dogs as well as all carnivores studied so far, but not in humans, a hyper-variable tandem repeat region is present in the control region (18–20). This repeat region exhibits a high amount of variation between individuals because of its very high rate of mutation (19). Because of the high mutation rate, it also shows heteroplasmy, i.e., more than one type of DNA is found in one individual (19,21,22). In dogs, the region is composed of an array of 10-bp sequences that are tandemly repeated, usually between 25 and 35 times (19). PCR amplification of this locus, therefore, results in a number of fragments of different lengths. Among dogs there are two types of these 10 bp repeats, differing in one position (informative position) (Fig. 1*a*). In some wolves, a third type has been found caused by variation in a second position (Fig. 1*b*). The different repeat types form a sequence and therefore there are two types of variation in this region, variation of length and variation of the sequence of repeat types. In a

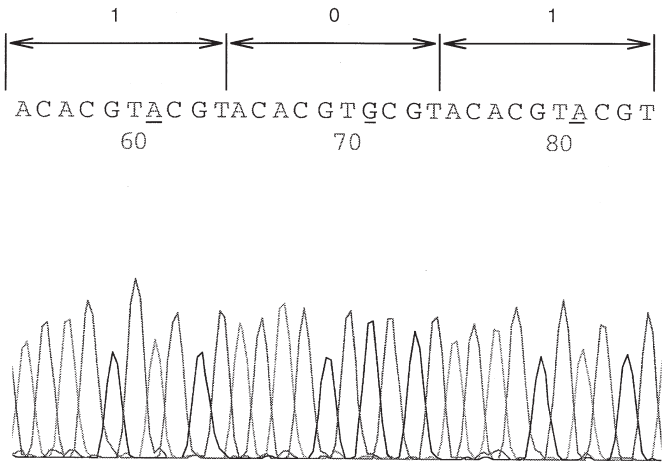
[1] Senior scientist and associate professor, respectively, Department of Biotechnology, KTH, Royal Institute of Technology, S-100 44 Stockholm, Sweden.

[2] Senior scientist, Department of Numerical Analysis and Computing Science, KTH, Royal Institute of Technology, S-100 44 Stockholm, Sweden.

* This work was supported by grants from the National Laboratory of Forensic Science (SKL), the National Board for Technical and Industrial Development (NUTEK) and the Swedish Research Council for Engineering Sciences (TFR).

Received 29 July 1999; and in revised form 5 Nov. 1999; accepted 8 Nov. 1999.

A



B

Repeat type	Sequence
0	ACACGT**G**CGT
1	ACACGT**A**CGT
2	ACAC**AT**ACGT

FIG. 1—(a) *Sequence analysis results showing three repeats. The informative position, which defines the two repeat types, 0 and 1, is underlined.* (b) *The three repeat types found in this study. The informative positions are written in bold.*

previous study, 15 dogs having mtDNA sequence variants representing all major phylogenetic mtDNA groups, according to analysis of the non-repetitive control region sequence, were analyzed (19). A very large amount of heteroplasmy was found in the individuals; 85 types of molecules (array types) were found in the blood within one individual and the total number of array types was estimated to about 380. In all individuals one or a few array types were more common and had frequencies between 10 and 50% while the majority of the array types had a frequency of less than 1%. In pedigree analyses, the more common array types were shown to be inherited from mother to offspring. The rare array types constituted in all approximately 20% of the molecules and were shown to be the product of somatic mutations. A pairwise comparison performed between all individuals in the study showed that while maternally related individuals had some identical array types, no inter-individual match was found between any two molecules from unrelated individuals.

This study investigates the possibility of utilizing the large amount of genetic variation found in the mtDNA tandem repeat region in dogs for forensic analysis. DNA extracted from blood and hairs was analyzed by fragment length analysis and DNA sequence analysis. Hereby, the possibility of using this region as a new type of forensic genetic marker was demonstrated.

Materials and Methods

Samples

Seventeen dogs and one wolf were sampled. The samples were chosen so that three non-repetitive mtDNA sequence variants of dogs (D5, D6, and D8) according to normal CR sequencing (15,17) were represented: seven individuals represented variant D5 (six German shepherds: H83, H9, H97, Ny68, Ny70, and Ny71, and one Dachshund: J1), five individuals and one pedigree represented variant D6 (one Golden retriever: M1, one Afghan hound: Ny56, one Buhund: Ny40, one Schipperke: Sch3, one four generation maternal pedigree of four Labrador retrievers [PD0, PD1, PD2, and PD3] and one Russian wolf: Wolf1) and two individuals represented variant D8 (one Jämthund: H54, one Norrbottenspets: Ny38). The samples were taken in the form of blood, hair or muscle according to the following list. H54, PD0, PD1, PD2, and PD3: blood samples. Sch3: hair and blood samples. Wolf: muscle sample. H9, H83, H97, J1, M1, Ny38, Ny40, Ny56, Ny68, Ny70, and Ny71: hair samples.

DNA Preparation

DNA was obtained from blood samples using the Chelex procedure (23). DNA was obtained from muscle samples using the following method: 1 cm^3 pieces were cut into slices and washed in 1 $\times$ SSC before the addition of 400 μL of 150 mM NaAc (pH 7.0), 1.25 mg/mL proteinase K, 50 mM DTT and 2% NP40 (non-ionic detergent). The samples were incubated at 37°C overnight and were then extracted twice with phenol/chloroform. DNA was recovered by two rounds of ethanol precipitation and dissolved in TE buffer. Hairs were sampled by pulling the hairs from the individuals so that the bulbs were recovered and DNA was obtained from the hairs using the following method: hairs were placed in a 1.5 mL tube containing 200 μL of hair digestion buffer: 10 mM Tris-HCl (pH 8.5), 0.9% Polyoxyethylene 10 Lauryl Ether, 35 mM DTT, 50 μg/mL proteinase K and 5% w/w chelex 100 (Bio-Rad, Richmond, CA). The mixture was incubated at 56°C overnight, at 96°C for 10 min and finally subjected to vortex mixing and centrifugation. The supernatant was used directly in the PCR amplification.

Direct DNA Amplification

The tandem-repeat containing region was amplified by PCR using the primers WD3: 5′-CAA GGT GCT ATT CAG TCA ATG G-3′ and WD6: 5′-TAT AAT AGA TGA CAT GAG TTT ACG-3′. For control experiments two more primers, situated internally of WD3 and WD6, were used; WD4: 5′-GGT TTG TAT AAG TTA ACT TAA TGT C-3′ and WD5: 5′-TTT CAG GAC ATA TAG TTT TAG GG-3′. For fragment length analysis WD3 or WD5 was fluorescently labeled with the 6-FAM dye label (PE Applied Biosystems, Foster City). The PCR mixture consisted of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 2 μg/mL BSA, 0.2 mM of each dNTP, 0.1 μM of each primer and 2 units of Ampli-*Taq* DNA polymerase (Perkin Elmer, Norwalk, CT) in a total volume of 50 μL. The amplification program consisted of a pre-denaturation step (94°C, 2 min) followed by 40 cycles of denaturation (94°C, 15 s), primer annealing (62°C, 30 s) and extension (72°C, 1 min) and a final extension step (72°C, 10 min).

DNA Cloning

Amplified DNA was ligated into the pGEM-T vector (Promega, Madison, WI) according to the manufacturer's directions and trans-

formed into Epicurian *Coli* competent cells (Stratagene, La Jolla, CA) and spread according to the manufacturer's directions, with the exception that 33 μL of cells instead of 100 μL were used for each sample.

DNA Amplification of Cloned DNA

Individual bacterial clones were picked and directly taken to PCR amplification of the cloned fragments. Plasmid specific primers; RIT27: 5′-GCT TCC GGC TCG TAT GTT GTG TG-3′ and RIT28: 5′-AAA GGG GGA TGT GCT GCA AGG CG-3′ were used. The PCR mixture was identical to that used for direct PCR. The amplification program consisted of a pre-denaturation step (94°C, 2 min) followed by 35 cycles of denaturation (94°C, 15 s), primer annealing (69°C, 30 s) and extension (72°C, 1 min) and a final extension step (72°C, 10 min).

DNA Fragment Analysis

Amplification product (0.2 μL) was mixed with deionized formamide and fluorescently labeled size standard; Tamra 500 (PE Applied Biosystems). The samples were analyzed on an ABI PRISM 377XL (PE Applied Biosystems) using 4% denaturing polyacrylamide gel according to the manufacturer's directions. Genescan software (PE Applied Biosystems) was used for size calling and quantification of the DNA fragments.

DNA Sequence Analysis

One μL of amplification product from cloned DNA was mixed with 3.2 pmole sequencing primer, USP: 5′-CGT TGT AAA ACG ACG GCC AG-3′, and BigDye cycle sequencing was performed according to the manufacturer's directions (PE Applied Biosystems). The cycle sequencing program consisted of 25 cycles of denaturation (96°C, 10 s), primer annealing (55°C, 5 s) and extension (60°C, 4 min). The sequencing reactions were ethanol precipitated and the pellets were air dried and dissolved in deionized formamide. DNA sequence analysis was performed on an ABI PRISM 377XL using 4% denaturing polyacrylamide gel according to the manufacturer's directions (PE Applied Biosystems). The resultant DNA sequences were studied for the presence of polymorphic positions using the SeqEd software (PE Applied Biosystems).

DNA Sequence Alignment and Pairwise Comparison

Sequence comparisons of array types were made by translating the DNA sequence to binary or trinary codes and aligning the codes using in-house developed software. The software tools for sequence comparisons on a Unix system is available on request from the authors.

Results

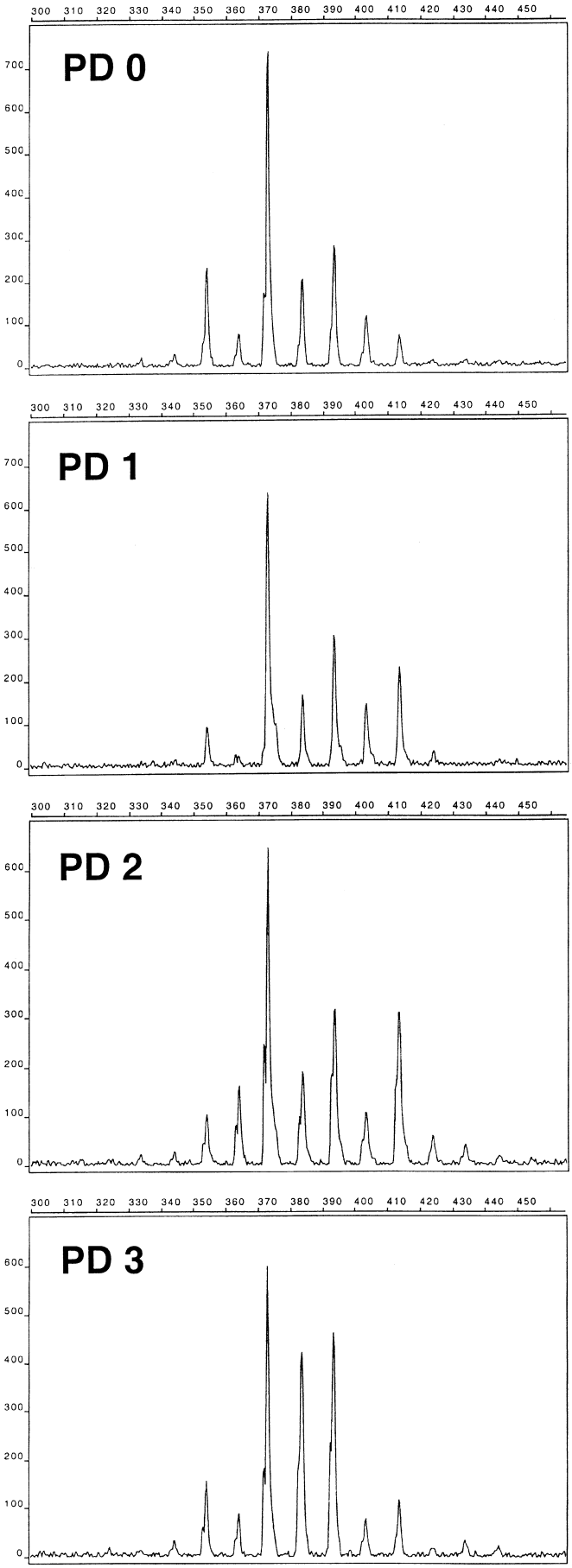
In order to evaluate the variability in the tandem repeat region of dog mtDNA, two experiments were performed: a, the length variation was studied by fragment length analysis, and b, the variation in the sequence of repeat types in individual molecules was studied by DNA sequence analysis.

Fragment Length Analysis

Total DNA was extracted from blood from four dogs in a four-generation maternal pedigree (PD0, PD1, PD2, and PD3), which, because of their relatedness, had identical non-repetitive mtDNA sequence variants. A PCR was performed using primers situated in non-repetitive regions on each side of the tandem repeat region. One of the primers was fluorescently labeled to enable automatic fragment length analysis. A large amount of heteroplasmy with seven to nine length variants was found in the individuals (Fig. 2). The height of the different peaks, i.e., the quantitative differences of length variants, fluctuated somewhat between the generations, with the result that the four individuals showed different heteroplasmic patterns, giving individualizing fragment length profiles. To make sure that the pattern was not PCR primer dependent, the same experiment was performed using another pair of primers, which were non-overlapping with the previous primers. This experiment gave the same fragment-length profiles (data not shown). To examine the stability of the heteroplasmic pattern over time in a single individual, blood was drawn from one animal, Sch3, at three different points of time, 0, 1, and 20 months, respectively. Analysis showed that consistent fragment analysis profiles were obtained from all three points of time (Fig. 3). However, when different hairs deriving from the same individual were analyzed, different fragment length patterns were obtained, even from hairs situated only millimeters apart (Fig. 4).

DNA Sequence Analysis of Cloned DNA

If the tandem repeat region is amplified from total DNA and directly sequenced, a mixed, non-interpretable sequence is obtained because of the heteroplasmic state of the individuals. Therefore, to be able to analyze the sequence of the different mtDNA molecules, it is necessary to clone the amplification product. To investigate the exclusion capacity of this type of analysis, several individuals representing the same non-repetitive sequence variant were compared. In a previous study, 19 non-repetitive mtDNA CR sequence variants were found among 102 domestic dogs (15). In this work, individuals representing three of these mtDNA sequence variants, called D5, D6, and D8 (15,17) were studied. The variants were found in 15, 9, and 4% of the individuals in the population, respectively (15). Seven individuals represented variant D5 (six German shepherds and one Dachshund), five individuals and one pedigree represented variant D6 (one Golden retriever, one Afghan hound, one Buhund, one Schipperke, one pedigree of four Labrador retrievers [PD0, PD1, PD2, and PD3], and one Russian wolf) and two individuals represented variant D8 (1 Jämthund, 1 Norrbottenspets). Between 3 and 218 clones from the different individuals were studied (Fig. 5) (data are shown for only one individual from the dog pedigree). Within the individuals from which several clones were analyzed, a large amount of heteroplasmy was found. In the individuals from which more than 20 clones were analyzed, at least, 10 array types were found. However, within most individuals, many identical molecules were found, and one or a few array types constituted together more than 50% of the molecules. An exception from this was individuals Ny56 and H54, from which 10 and 20 molecules were analyzed, respectively, which were all different. In order to examine the genetic variation among individuals, a pairwise comparison of repeat-type sequences between individuals was performed on all possible pairs of clones. Within the pedigree, several array types were found in clones from all four individuals (data not shown), but among unrelated individuals, no array type matches were found, not even between individuals sharing the same non-repetitive CR sequence variant. By looking at the array types, it is obvious that generally there is a larger difference between than within individuals (Figs. 5 and 6). However, some individuals have more similar array types. For example, all German



shepherds have identical 3′-ends in most of their array types, and three of them have majority array types with an identical sequence of 16 repeats in the 3′-end. Furthermore, four of the six individuals of variant D6 have an identical sequence of seven repeats in the 3′-ends.

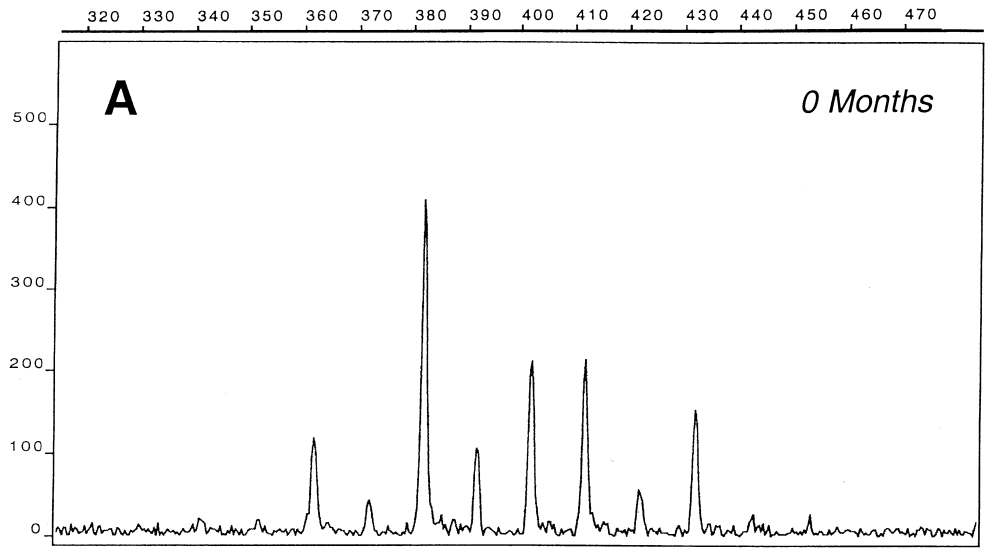
To investigate whether a blood sample and different hairs taken from one individual contain the same array types and can be linked to each other, one blood sample and three hairs (the corresponding fragment analyses, hairs II to IV, are shown in Fig. 4), were collected from Sch3 and analyzed. Thirty-two clones were sequenced from the blood sample and from the three hairs 20, 22, and 25 clones were sequenced (Fig. 7). A slightly higher amount of genetic variation was found in the blood sample than in the hair samples. Among the 32 clones from blood 25 array types were found, while the hair samples (20, 22, and 25 clones) resulted in 9, 12, and 16 variants, respectively. However, the important facts are that eight different array types were shared by two or more of the samples and all four samples shared two array types (numbers 10 and 17). Array type 10 was the most common in all samples, representing between 25 and 36% of the molecules. Thus, despite the variation between the hairs found in the fragment analyses, a genetic link can be obtained by the analysis of repeat-type sequence. To ascertain that the results are stable in time, a blood sample drawn 20 months later than the first one was analyzed. Eight clones were analyzed, and two of these were of the majority type from the first blood sample (data not shown). Furthermore, the fact that identical array types are found in maternally related individuals also shows that the repeat-type sequences of the array types are stable in time.

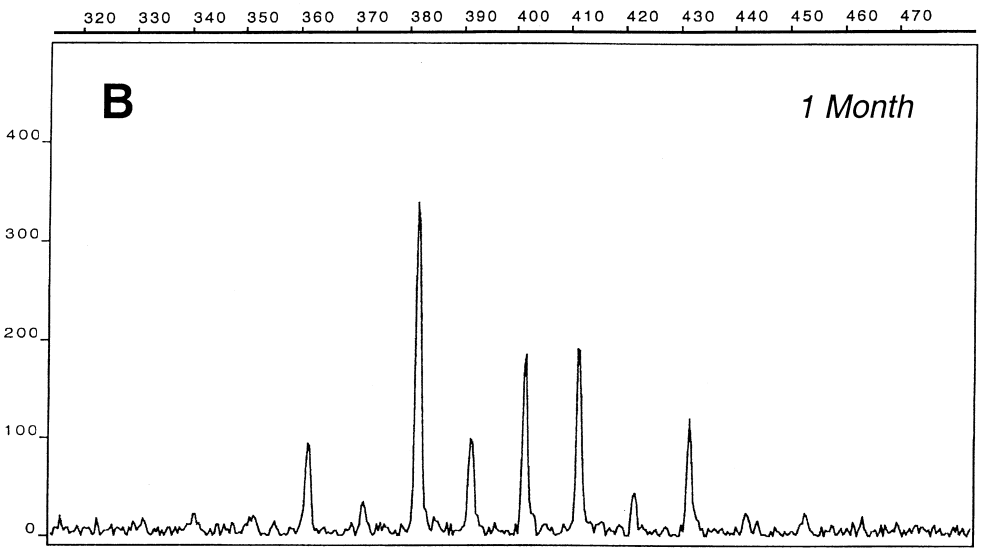
Discussion

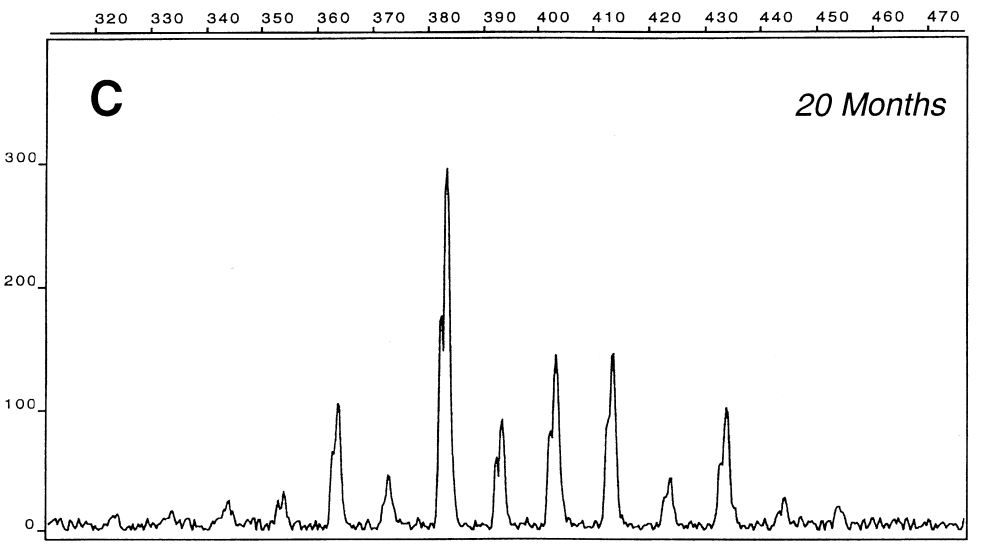
We have previously demonstrated the use of analysis of mtDNA from dog hairs in forensic investigations of murder, theft, and poaching (14). In these studies, the sequencing of nonrepetitive mtDNA has been used. It has proved to be a useful method in several cases, especially for the linking of different crimes and for the exclusion of suspects. However, in the case that an individual is included as a possible source of the sample, it can often be used only as weak circumstantial evidence, because of the high frequencies of some mtDNA sequence variants. Therefore, in order to investigate the possibility of obtaining a higher exclusion capacity for the analysis of dog mtDNA, the tandem repeat region was studied.

The study indicates that fragment-length analysis of this region can be used to distinguish between individuals even within the same family when DNA is obtained from blood. The fluctuating proportions of the length variants in the pedigree can probably be explained by a bottleneck in the female germline, resulting in random drift among the molecules (24). Analysis of blood samples from one individual at several different points of time showed stable amounts of the different length variants. However, when DNA was obtained from single hairs, a large difference in the fragment-length pattern was seen between different hairs from one and the same individual. The method is therefore useful for the comparison of bloodstains and blood samples from dogs, but not for the

FIG. 2—*Fragment analysis of the tandem repeat region. PCR products from blood samples taken from individuals PD0, PD1, PD2, and PD3, which represent a straight four-generation maternal line. Length of fragments in base-pairs and relative intensity of fluorescence are shown on the X and Y axis, respectively.*


320 330 340 350 360 370 380 390 400 410 420 430 440 450 460 470
A
0 Months
500
400
300
200
100
0


320 330 340 350 360 370 380 390 400 410 420 430 440 450 460 470
B
1 Month
400
300
200
100
0


320 330 340 350 360 370 380 390 400 410 420 430 440 450 460 470
C
20 Months
300
200
100
0

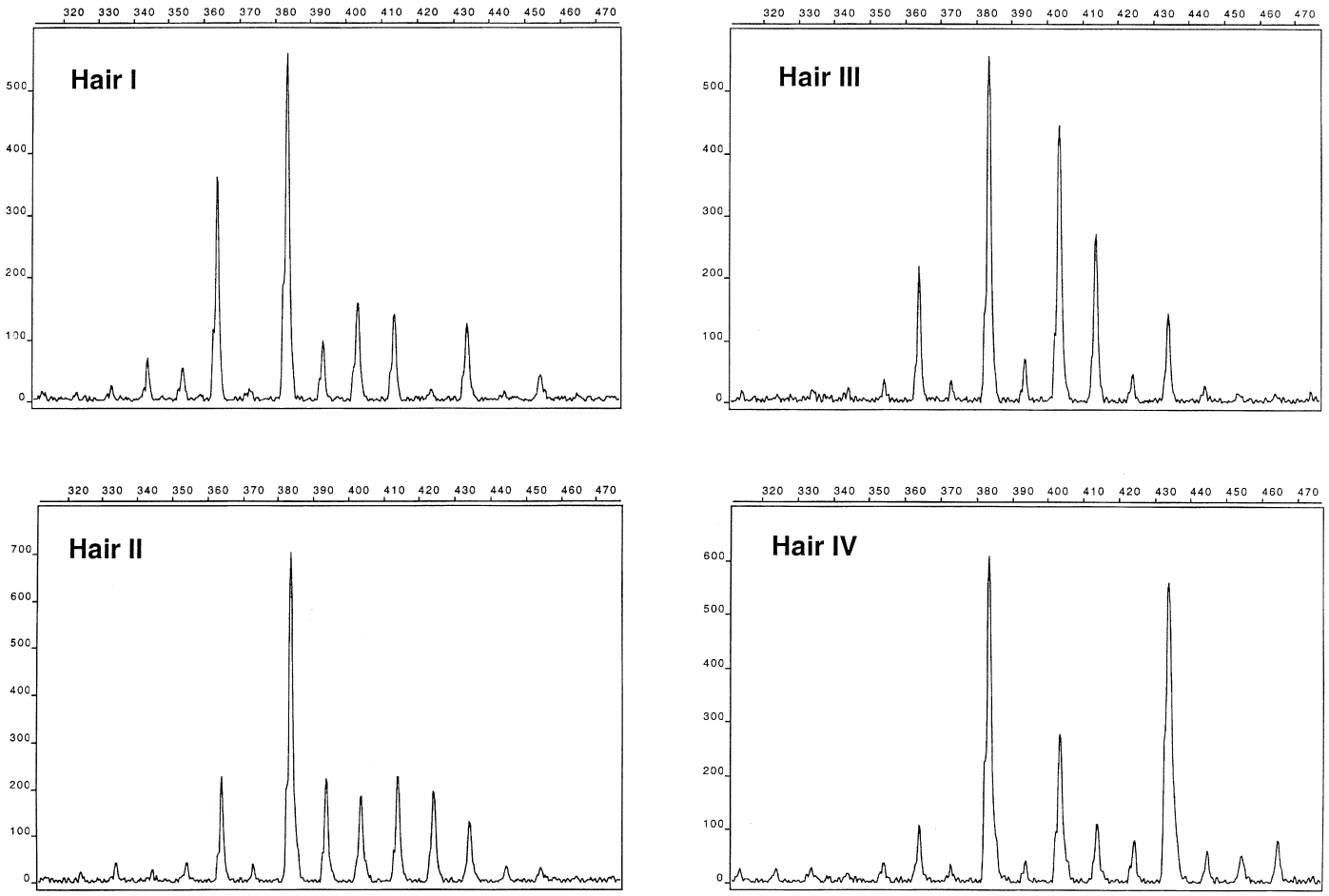


FIG. 4—*Fragment analysis of the tandem repeat region. PCR products from four hairs of one individual, Sch3. Hairs 1 and 2 are sampled within 1 cm^2 on the left foreleg. Hairs 3 and 4 are sampled within 1 cm^2 on the right hindleg. Length of fragments in base-pairs and relative intensity of fluorescence are shown on the X and Y axis, respectively.*

comparison of single hairs to each other or to blood samples. Different proportions of mtDNA types in different hairs have also been reported in humans that are heteroplasmic for point mutations (25,26). The fact that the hairs show a much larger variation in the heteroplasmic pattern compared with the blood samples may be explained by the fact that each hair stems from a small number of stem cells in the hair follicle, resulting in a high degree of random drift among the molecules. In contrast, peripheral blood consists of lymphocytes produced from very large numbers of hematopoietic stem cells (25). Interestingly, when the peak heights of the four hairs are summed up, the resulting relative values of the peak heights are similar to those of the blood sample (data not shown).

However, studies of the array types showed that two blood samples and three hairs contained common array types. Therefore, if only a sufficient number of molecules are analyzed from a hair, it should always be possible to find array types that are common with those in a blood sample. Therefore, the analysis of repeat type-sequence seems to be a suitable method for hair investigations.

When a pairwise comparison of all array types was performed between the 15 unrelated individuals, representing three non-repetitive mtDNA sequence variants, no two identical array types were found. Thus, the tandem-repeat region displays a large amount of variation between individuals. From a few individuals, only a small number of clones were analyzed, with the result that only unique array types were found. It could therefore be possible that array types identical to those of other individuals have been overlooked. However, this does not seem probable, as the array types found in these individuals differ considerably from those of the other individuals.

In order to obtain an estimate of the exclusion capacity of the region, a very conservative estimate was done of the amount of genetic variation of the 16 other non-repetitive sequence variants

FIG. 3—*Fragment analysis of the tandem repeat region. PCR products from three blood samples taken from one individual, Sch3, at 0, 1 and 20 months. Length of fragments in base-pairs and relative intensity of fluorescence are shown on the X and Y axis, respectively.*

Ind.	A.t.	N	Repeat type sequence
H9	24	3	11110010110011001010100000 1001011
(D5)	1	1	11101110011011010101010000 1001011
	5	2	11110010110010010101000001 001011
	8	1	11100110110101010101000001 001011
	15	1	11110101100100101010000 1001011
	16	2	11110011011010101010000 1001011
	39	2	11101101101010101000010 1001011
	11	1	11100110110101010100001 0001011
	10	16	11100110110101010100001 0001011
	18	2	11001101101010101000010 0001011
H83	31	1	11111100101010101010101010001 001011
(D5)	32	1	11101001000101011101010100010 001011
	40	2	11101001000101010101010001 001011
	36	2	11100100010101010101010001 001011
	43	2	11110010001010101010101001 001011
	28	1	11100100010101010101010001 001011
	37	12	11100100010101010101010001 001011
	17	1	11100100010101010101000001 001010
	42	1	11101001000101010101010001 011
	9	1	11101000101010101010001 001011
	47	1	10000000010100100000000001 011
	24	1	11000101010101010001 001011
	27	1	1110010001010100 1011
H97	11	1	00101010101010100001010010000000 11
(D5)	27	1	11110000101010000101001000000 11
	36	1	10010101010100001010010000000 11
	40	7	10010101010100001010010000000 11
	39	1	10000101010100001010010000000 11
	22	1	11000010101000010100100000 11
	13	1	01001010100001010010000000 11
	10	4	10010101010000101001000000 11
	35	2	10001010100001010100100000 11
	19	1	11001010100001010010000000 11
	14	2	00101010100001010010000000 11
	32	1	10010101000010100100000 11
	4	1	10010010010100100000000 1011
	5	2	00010000001010010000001 011
	31	1	10010101010000101000000 11
	12	1	10010101000010100100000 11
	1	8	10010010010100100000001 011
	9	1	11110000101010000 1001011
Ny68	34	1	11100110101100101001010100001 001011
(D5)	35	1	11100110101010101000000 10010111
	6	1	11100100101010101000001 001011
	5	1	11100101000101010100001 001011
	2	3	11100110101010100000 1001011
	19	1	11100101001010101000001 001011
	12	5	11111001001010101000 1001011
	4	1	11100110101010101000 1001011
	30	1	11100110101010100000 1001011
	4	1	11100100101010100000 1001011
	36	1	11001001010101000001 001011
	31	1	11100100101010100 1001011
	29	1	11100100101010000 1001011
	28	1	10010101101000001 001011
	16	1	11100100100 1011
Ny70	17	1	10101010101010000010101010001 001011
(D5)	16	1	10101001010000000010101000 1001011
	21	1	10101010100000100101010001 001011
	8	1	10001010100000100101010001 001011
	20	1	10101011000001001010100 1001011
	4	1	10101010100001010100 1001011
	2	2	10101010100001010100 1001011
	7	1	10110010101000001000 1001011
	18	1	10101010000101010001 001011
	14	1	10101010100010100 1001011
	1	2	10101000010101000 1001011
J1	7	1	11101010110101010100101010001
(D5)	14	1	11010101011010101010 1010001
	6	1	11010101011010101010000 10001
	9	3	11010101101010101001010001
	17	1	11000101101010101001010001
	16	1	11110101011010100 1010001
	15	1	11010101010010100 1

Ind.	A.t.	N	Repeat type sequence
Ny71	16	9	11010111010001010101000 0010111
(D5)	3	1	01011101010001010101000 0010111
	31	1	10101110100010101010000 0010111
	30	1	01101110100010101010000 0010111
	19	1	01001110100010101010000 0010111
	20	1	11010111010001010100000 1010111
	38	4	01011101000101010100000 1010111
	44	1	01011101000010101000001 0010111
	1	1	11110100010101010000 1010111
	13	3	01011101000101010000 1010111
	15	1	01011101000010101000 1010111
	4	1	11101000101010100000 1010111
	37	1	01010001010101000001 010111
	36	1	11010101010100000 1010111
M1	13	1	11001110011001000011100110010001
(D6)	18	1	1100111001100100001110101101101
	3	1	1111000111001000011101010 1101
	5	1	1000110011001000011101010 1101
	11	1	1100110011001000011101010 1101
	2	1	1100111101100100001010 1101
Wolf1	59	1	12201021200000212121201100001 100000001
(D6)	47	1	12212000121212120201100011 0100000001
	4	1	12200121202011000110000001 100000001
	i1w	1	12212000121212120201100011 00000001
	1	1	12211000121212120201100011 00000001
	6	1	12221200121202011000110000000001
	5	1	12212000121212120200001100000001
	29	1	12122120001212020110001100000001
	2	1	12212001212120201100001000000001
	i3w	1	12200121212120201100011 00000001
	13	1	12200121212020110100011 00000001
	i2w	3	12200101212120201100011 00000001
	63	1	21200012120201100011000000001
	14	1	12212000121212120200110000001
	58	1	12200121212122011100011 0000001
	10	1	12200121212020110001000000001
	15	1	11220012121202011000110000001
	23	1	12201220012121202011001100001
	i6w	31	12200121212020110001100000001
	18	2	12200101212120110001100000001
	68	1	12222120012121202011000000001
	27	1	12200121212020221200012212021
	31	1	12200121212020110012000000001
	39	1	12200121212011000110000000001
	43	1	11212120201100011000000000001
	25	1	12212121202011000110000000001
	61	1	12022120201100011000000001
	i9w	1	1220201100011000000001
	7	1	1220110100000001
Sch3	7b	1	10101101010010100100101001101011001
(D6)	3b	1	10100101001000010100101101011010 1001
	18	1	11010100101010010010100101101011001
	8	1	10101001010010010100101101101011001
	17	2	10101101010010010010011010 1101
	3	1	10101010010100100101001011010 1001
	24	1	10101001010010010100101101011001
	4	1	10101001010010010100101101001001
	19	1	10010100100100001010011010 1101
	21	1	01001001101001001010010110101101
	13	1	01001001010011010100101101011101
	7	1	10101101010010100100100101 1001
	1	1	10100101001001001010011010 1001
	8b	1	10100101001000010100101011010 1
	23	1	10010101001000010101011010 1101
	15	1	10010100100001010010110101 1001
	10	8	10101001010010010010110101101
	14	2	10010100100001010011010 1101
	6b	1	10100101001000010101101011010 1
	12	1	10000101001000010100101111010 1
	9	1	10100101001001001010110101101
	5b	1	00101001001010010011010 1101
	4b	1	111010111101
		0	
		0	

FIG. 5—*Array types from the compared individuals. The dog pedigree is represented by PD1. Individual (non-repetitive mtDNA sequence variant is given within brackets), array type, number of clones and repeat-type sequence are given. The array types from PD1 that were found in one or several of the other members of the pedigree are indicated by the prefix "i" in the array type names.*

Ind.	A.t.	*N*	Repeat type sequence
PD1	i2	1	11111111111001000011101010101101
(D6)	8a	1	11111111100100000111010110101101
	49a	1	11111111100100001110101101101101
	i4	5	11111111001000001110101010110101
	29a	1	11111100100000001110101010110101
	i5	1	11111111001000011101010101101
	i6	1	11111110010000011101010101101
	13a	1	11111100100000011101010101101
	i10	1	11111111001000011101101101101
	i9	2	11111110010000111010101101101
	i8	4	11111100100000111010101101101
	40a	1	11111001000011101010101101101
	46a	1	11111111000011101010110101
	i12	1	11111110010000111011010101101
	i11	1	11111100100001110101010101101
	i16	3	11111100100001110101101101
	i15	5	11111001000011101010101101
	i17	2	11110010000011101010110101
	23a	1	11110010000111010101101101
	i18	1	11110010000111010101101101
	42a	1	11100100000111010101101101
	27a	2	11111110010000101011011101
	i20	1	11110010000111011010101101
	i19	3	11100100001110101010101101
	41a	1	1100000111010101011101
Ny40	15	1	11110101010110101010010100101101
(D6)	14	1	1110010111010101001010010101101
	12	1	10010101000101101
Ny56	8	1	10011101011010001110010101000011100000001
(D6)	6	1	11111101001100101101000011100000001
	10	1	10101101101010101101101010101011
	1	1	11111010011001010100001110000001
	9	1	11111010100101011000011100000001
	5	1	11110100110010101000011101000001
	4	1	11110100110010101000011100000001
	3	1	1111010011001010100001111010001
	7	1	10011101011010001110000000001
	2	1	11110100110010101010101011
H54	21	1	11100000001110000000011001110000011001
(D8)	10b	1	11111110100000000110011100011001
	6b	1	11101111010000000110011100011001
	11b	1	11111110100000001100111000011001
	3b	1	11110101000000001100111000011001
	18	1	11101110100000111001110001001
	4b	1	11101010110000011100111000011001
	8b	1	11111010000000011001110001001
	14	1	11110101100000111001110001001
	1b	1	11101101100000111001110001001
	1	1	1110101100000111001110001001
	22	1	11101011000001110000000001101
	17	1	11101000000011001110000011001
	9	1	1110110000011100111000011001
	9b	1	11101000000011001110011001
	7b	1	11100000000110011100011001
	5b	1	11000000000110011100011001
	2b	1	1101000000001100111001
Ny38	2	1	11101000100010010100000100001
(D8)	4	1	1110100100100010010000000100001
	3	1	111010100010010010000000100001
	1	1	11101000100100100000100001

FIG. 5—*(Continued.)*

Nonrep. seq.var.	Breed	Ind.	Array type	Repeat-type sequence
5	G.Shepherd	H9	10	11100110110101010100001001011
5	G.Shepherd	H83	37	11100100010101010101000100101l
5	G.Shepherd	H97	1	10010010010100100000001011
5	G.Shepherd	Ny68	12	11111001001010100010010l1
5	G.Shepherd	Ny70	2	10101010100001010100010010l1
5	G.Shepherd	Ny71	16	11010111010001010101000010001011
5	Dachshund	J1	9	11010101101010101001010001
6	G.Retriever	M1	3	11110001100100001110101101
6	L.Retriever	PD1	i15	11111001000011101010101101
6	Schipperke	Sch3	10	10101001010010010010110101101
6	Buhund	Ny40	14	11100101110101010010100101101
6	Afghan	Ny56	4	11110100110010101000011100000001
6	Wolf	Wolf1	i6w	12200121212020110001100000001
8	Jämthund	H54	1	11101011000001110011100011001
8	Norrb.spets	Ny38	3	11101010001001001000000100001

FIG. 6—*A comparison of array types from the 15 individuals studied. The most frequent array type or, in the case of equal frequencies, an arbitrarily chosen array type is shown for the 15 individuals. Nonrepetitive sequence variant, breed, individual, array type, and repeat-type sequence are given.*

previously found (15) but not studied here. This was based on the fact that all six individuals of sequence variant D6 had different array types and that the two individuals of variant D8 had different array types. The sequence types with a frequency equal to or higher than that of variant D6 (9%) were therefore divided into six equally frequent groups, and the sequence variants with a frequency higher than that of variant D8 (4%) were divided into two equally frequent groups. This gives a total of 48 different groups having different array types among the 102 dogs in the population, and a total exclusion capacity (defined as 1 minus the sum of the squares of the frequencies of the groups) of 0.976 for this genetic marker. This figure is comparable to that of the control region in humans. However, as all six German shepherds of sequence variant D5 can be distinguished by the analysis, and as it can be assumed that some of these individuals have a relatively recent common maternal ancestor, it is probable that all dogs of different breeds having the same sequence variant can be distinguished by their array type. This would imply at least 74 groups of array types among the 102 dogs studied. In conclusion, this study shows that the analysis of array types in dogs adds a substantial amount of exclusion capacity to forensic mtDNA analysis of domestic dogs.

The facts that all six German shepherds could be distinguished and that the breed was defined and breeding was started as late as 1899 (27) indicate a very high evolutionary rate in the repeat-type sequence. Pedigree analysis could be a possible means to the studying of the divergence rate and mode of evolution of this region. This, in combination with the study of a larger population of individuals, would make it possible to evaluate its full value in terms of exclusion capacity. The drawbacks of this method are that the PCR product must be cloned and that several clones, probably at least 20, must be sequenced from each sample in order to enable an evaluation of the genetic variation within the individual. Both these facts imply that the analysis method is more laborious than normal DNA sequence analysis. However, due to its seemingly very high exclusion capacity, it can have a great value in investigations of serious crimes. The same type of imperfect repeats is also found in other carnivores, and especially in the case of domestic cats forensic analysis of this region could be valuable.

In conclusion, this study shows that the tandem repeat locus in dog and wolf mtDNA is highly divergent between individuals, because of variation in the sequence of repeat types. Furthermore, DNA from single hairs and blood samples from one individual had identical array types and stable results were obtained from samples collected at an interval of 20 months. The method therefore seems suitable for analysis of DNA from single hairs for the

Sample	Array type	Frequency	Repeat type sequence
Blood	7b	1	10101101010010100100101001011011001
	3b	1	10100101001000010100101101011001001
	18	1	11010100101010010010100101101011001
	8	1	10101001010010010100101101101011001
	17	1	10101101010010100100100101101101
	2b	1	1010110101001010010010010110101101
	3	1	10101010010100100101001011010011001
	24	1	101010010100100101001011010011001
	4	1	1010100101001001010010110100 1001
	19	1	1001010010010000101001011010 1101
	21	1	0100100110100100101001011010 1101
	13	1	010010010100110101001011010 1101
	7	1	101011010100101001001001011001
	1	1	101001010010010100101101011001
	8b	1	101001010010000101001010110101
	23	1	100101010010000101010110101101
	15	1	100101001000010100101101011001
	10	8	101010010100100100101101101
	14	2	100101001000010100101101101
	6b	1	10100101001000010101101101101
	12	1	10000101001000010100101111101
	9	1	101001010010010010110101101
	5b	1	00101001001010010110101101
	4b	1	111010111101
Hair I	2	1	10110101001010010010011010101101
	13	5	10101101010010100100100101101101
	42	1	10101100101010010010100101101101
	43	1	10011010100101001001001011010 1101
	45	1	10110101001010010010010110101101
	6	1	10010000101001001010010110101101
	22	*3*	*10101101010010010010110101101*
	10	6	101010010100100100101101101
	41	1	010100100101001011010 1101
Hair II	37	1	10101011010100101001001001011010 1101
	17	2	101011010100101001001001011010 1101
	27	1	10101101010010100100100110101101
	35	1	10100101001010010010010110101101
	16	1	10101010010100100100101101101101
	30	*1*	*10100101001000010100101101101101*
	39	*1*	*01001001010010010100101101101101*
	10	8	101010010100100100101101101101
	32	1	10101001010010010010100101101
	14	2	100101001000010100101101101101
	9	2	101001010010010010110101101
	28	1	100101001001010010110101101
Hair III	34	1	10101001010001010010010010110101101
	17	1	101011010100101001001001011010 1101
	36	1	10101101001010010001001011010 1101
	24	3	101010010100100101001011010 11001
	46	1	10101101010010001001001011011 0101
	30	*1*	*10100101001000010100101101101101*
	33	1	10010101001000010100101101101101
	39	*1*	*01001001010010010100101101101101*
	5	1	11010110101001010010010011011 0101
	22	*1*	*10101101010010010010110101101*
	10	7	101010010100100100101101101101
	14	2	100101001000010100101101101101
	44	1	10101001010010100101101101101
	40	1	10101001000100100101101101101
	9	1	101001010010010010110101101
	11	1	10100101001010010110101101 1001

discrimination between individuals, and represents a new approach to forensic DNA analysis: mtDNA repeat-type sequence analysis.

References

1. Jeffreys AJ, Royle NJ, Patel I, Armour JA, MacLeod A, Collick A, et al. Principles and recent advances in human DNA fingerprinting. EXS 1991;58:1–19.
2. Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 1989;44:388–96.
3. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 1991;49:746–56.
4. Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Methods Appl 1993;3:13–22.
5. Jeffreys AJ, Wilson V, Thein SL. Hypervariable "minisatellite" regions in human DNA. Nature 1985;314:67–73.
6. Jeffreys AJ, Wilson V, Thein SL. Individual-specific "fingerprints" of human DNA. Nature 1985;316:76–9.
7. Nass MM. Mitochondrial DNA. I. Intramitochondrial distribution and structural relations of single- and double-length circular DNA. J Mol Biol 1969;42:521–8.
8. Bogenhagen D, Clayton DA. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. J Biol Chem 1974; 249:7991–5.
9. Hopgood R, Sullivan KM, Gill P. Strategies for automated sequencing of human mitochondrial DNA directly from PCR products. Biotechniques 1992;13:82–92.
10. Sullivan KM, Hopgood R, Gill P. Identification of human remains by amplification and automated sequencing of mitochondrial DNA. Int J Legal Med 1992;105:83–6.
11. Allen M, Engstrom AS, Meyers S, Handt O, Saldeen T, von Haeseler A, et al. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. J Forensic Sci 1998;43: 453–64.
12. Tully G, Sullivan KM, Nixon P, Stones RE, Gill P. Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. Genomics 1996;34:107–13.
13. Menotti-Raymond MA, David VA, O'Brien SJ. Pet cat hair implicates murder suspect. Nature 1997;386:774.
14. Savolainen P, Lundeberg J. Forensic evidence based on mtDNA from dog and wolf hairs. J Forensic Sci 1999;44:77–81.
15. Savolainen P, Rosen B, Holmberg A, Leitner T, Uhlen M, Lundeberg J. Sequence analysis of domestic dog mitochondrial DNA for forensic use. J Forensic Sci 1997;42:593–600.
16. Piercy R, Sullivan KM, Benson N, Gill P. The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Legal Med 1993;106:85–90.
17. Vila C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, et al. Multiple and ancient origins of the domestic dog. Science 1997;276:1687–9.
18. Hoelzel AR, Lopez JV, Dover GA, O'Brien SJ. Rapid evolution of a heteroplasmic repetitive sequence in the mitochondrial DNA control region of carnivores. J Mol Evol 1994;39:191–9.
19. Savolainen P, Arvestad L, Lundeberg J. mtDNA tandem repeats in domestic dogs and wolves; mutation mechanism studied by analysis of the sequence of imperfect repeats. Mol Biol Evol. In press.
20. Lopez JV, Cevario S, O'Brien SJ. Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. Genomics 1996;33:229–46.
21. Hoelzel AR, Hancock JM, Dover GA. Generation of VNTRs and heteroplasmy by sequence turnover in the mitochondrial control region of two elephant seal species. J Mol Evol 1993;37:190–7.
22. Xu X, Arnason U. The complete mitochondrial DNA sequence of the horse, Equus caballus: extensive heteroplasmy of the control region. Gene 1994;148:357–62.
23. Walsh PS, Metzger DA, Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 1991;10:506–13.
24. Jenuth JP, Peterson AC, Fu K, Shoubridge EA. Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nat Genet 1996;14:146–51.
25. Bendall KE, Macaulay VA, Sykes BC. Variable levels of a heteroplasmic point mutation in individual hair roots. Am J Hum Genet 1997; 61:1303–8.
26. Wilson MR, Polanskey D, Replogle J, DiZinno JA, Budowle B. A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes. Hum Genet 1997;100:167–71.
27. Forster U. Der Deutsche Schaferhund. Falken Verlag, 1993.

Additional information and reprint requests:
Joakim Lundeberg, Ph.D.
Dept. of Biotechnology
KTH, Royal Institute of Technology
S-100 44 Stockholm
Sweden

FIG. 7—*Array types obtained from blood and hairs from one individual, Sch3. Shared array types have the same array type identification in the different samples. Shared array types present in the blood sample are underlined. Array types shared between hairs only are in italics.*