

# mtDNA Tandem Repeats in Domestic Dogs and Wolves: Mutation Mechanism Studied by Analysis of the Sequence of Imperfect Repeats

Peter Savolainen,\* Lars Arvestad,† and Joakim Lundeberg\*

\*Department of Biotechnology and †Department of Numerical Analysis and Computing Science, Royal Institute of Technology (KTH), Stockholm, Sweden

The mitochondrial (mt) DNA control region (CR) of dogs and wolves contains an array of imperfect 10 bp tandem repeats. This region was studied for 14 domestic dogs representing the four major phylogenetic groups of nonrepetitive CR and for 5 wolves. Three repeat types were found among these individuals, distributed so that different sequences of the repeat types were formed in different molecules. This enabled a detailed study of the arrays and of the mutation events that they undergo. Extensive heteroplasmy was observed in all individuals; 85 different array types were found in one individual, and the total number of types was estimated at 384. Among unrelated individuals, no identical molecules were found, indicating a high rate of evolution of the region. By performing a pedigree analysis, array types which had been inherited from mother to offspring and array types which were the result of somatic mutations, respectively, could be identified, showing that about 20% of the molecules within an individual had somatic mutations. By direct pairwise comparison of the mutated and the original array types, the physiognomy of the inserted or deleted elements (indels) and the approximate positions of the mutations could be determined. All mutations could be explained by replication slippage or point mutations. The majority of the indels were 1–5 repeats long, but deletions of up to 17 repeats were found. Mutations were found in all parts of the arrays, but at a higher frequency in the 5' end. Furthermore, the inherited array types within the mother-offspring pair were aligned and compared so that germ line mutations could be studied. The pattern of the germ line mutations was approximately the same as that of the somatic mutations.

## Introduction

Extensive size variation of the mtDNA control region (CR), caused by variation in the number of tandemly repeated sequences, has been described for a number of species. In vertebrates, arrays of repeats are found in five positions, referred to as repetitive sequence (RS) 1–5 (Hoelzel et al. 1994), all situated in regions where H-strand replication is regulated (fig. 1). RS1 and RS2 are found in the 5' end of the CR where the H-strand replication pauses, forming the three-stranded D-loop. RS3, RS4, and RS5 are situated in the 3' end of the CR, upstream of the site for H-strand origin of replication, and, consequently, near the end of the same replication (Clayton 1991; Hoelzel et al. 1994). A high degree of internal complementarity and the capability to fold into complex secondary structures have been described for repeat arrays at all five positions. The arrays at RS1, RS2, RS4, and RS5 have repeats of 40–160 bp showing variation in the number of tandem repeats and, often, length heteroplasmy. The arrays at RS3 are distinctive compared with those at RS1, RS2, RS4, and RS5 in that the repeats are shorter (6–22 bp), the numbers of repeats are higher (up to >40), and the level of heteroplasmy is higher. Tandem repeat arrays at RS3 were found in all 18 carnivores studied (Hoelzel et al. 1994), including domestic dog, and in 14 shrew species (Fumagalli et al. 1996) and have been found in several other mammals, such as pigs (Ghivizzani et al. 1993) and rabbits (Mignotte et al. 1990). In the carnivore, the arrays were composed of imperfect repeats forming a secondary sequence of repeat types which differed be-

tween molecules within individuals, but in pigs and rabbits, the arrays were composed of perfect repeats.

Although many studies have been performed on the repeat arrays of the CR, little is known about the mechanisms governing the length mutations, the maintenance of the number of repeats in the arrays, and, in the case of imperfect repeats, the apparent homogenization of the sequence of the repeats within species. Several mechanisms have been proposed to be the cause of the variation in copy number of the repeats. The most commonly suggested mechanism is slipped-strand mispairing during replication, also known as replication slippage (Levinson and Gutman 1987; Buroker et al. 1990; Fumagalli et al. 1996; Broughton and Dowling 1997; Wilkinson et al. 1997), but recombination and transposition (Rand and Harrison 1989) and unequal crossing-over or gene conversion (Hoelzel, Hancock, and Dover 1993; Hoelzel et al. 1994) have also been proposed to fit with various data. However, the methods used in these studies have not been able to directly record the mutational events.

In this study, the tandem-repeat region of domestic dog and wolf CR mtDNA was studied. As the repeat arrays of this region are composed of imperfect repeats, it is possible to establish the physiognomy of inserted or deleted elements and their approximate positions. By studying the inheritance of mtDNA molecules from mother to offspring, somatic mutations were recorded by direct comparison between the original and the mutated molecules. In combination with an analysis of the heteroplasmic variation in the domestic dog population, a comprehensive picture of the mechanisms involved in the evolution of the tandem-repeat arrays was obtained.

## Materials and Methods

### Samples

Fourteen dogs and 5 wolves were sampled. The dogs were chosen so that the four major phylogenetic

Key words: mitochondrial DNA, tandem repeats, mutation, *Canis*.

Address for correspondence and reprints: Joakim Lundeberg, Department of Biotechnology, Royal Institute of Technology (KTH), S-100 44 Stockholm, Sweden. E-mail: joakim.lundeberg@biochem.kth.se.

*Mol. Biol. Evol.* 17(4):474–488. 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

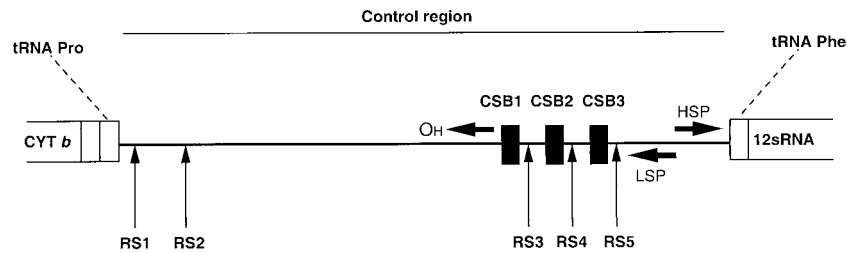


FIG. 1.—Schematic diagram of the organization of the mitochondrial DNA control region in mammals. The conserved sequence blocks (CSBs), the H-strand replication origin ( $O_H$ ), the H- and L-strand promoters (HSP and LSP, respectively), and the locations of tandemly repeated sequences in different species (RS1–RS5) are shown (Walberg and Clayton 1981; Clayton 1991; Saccone, Pesole, and Sbisà 1991; Hoelzel et al. 1994).

groups of dogs according to normal CR sequence were represented (Savolainen et al. 1997; Vila et al. 1997). Three sequence variants were represented by three (variant D5), three (variant D6), and two (variant D8) unrelated individuals, respectively, and five variants were represented by a single individual. Furthermore, two three-generation maternal pedigrees were studied, one

consisting of dogs and one consisting of wolves. Both pedigrees were of sequence variant D6 (also called W6 when found in wolves). Studied individuals were as follows: H9 (German shepherd, sequence variant D5); H54 (Jämthund, D8); H83 (German shepherd, D5); PD1, PD2, and PD3 (maternal pedigree, Labrador retriever, D6); Ny34 (border collie, D1); Ny36 (Samoyed, D7);

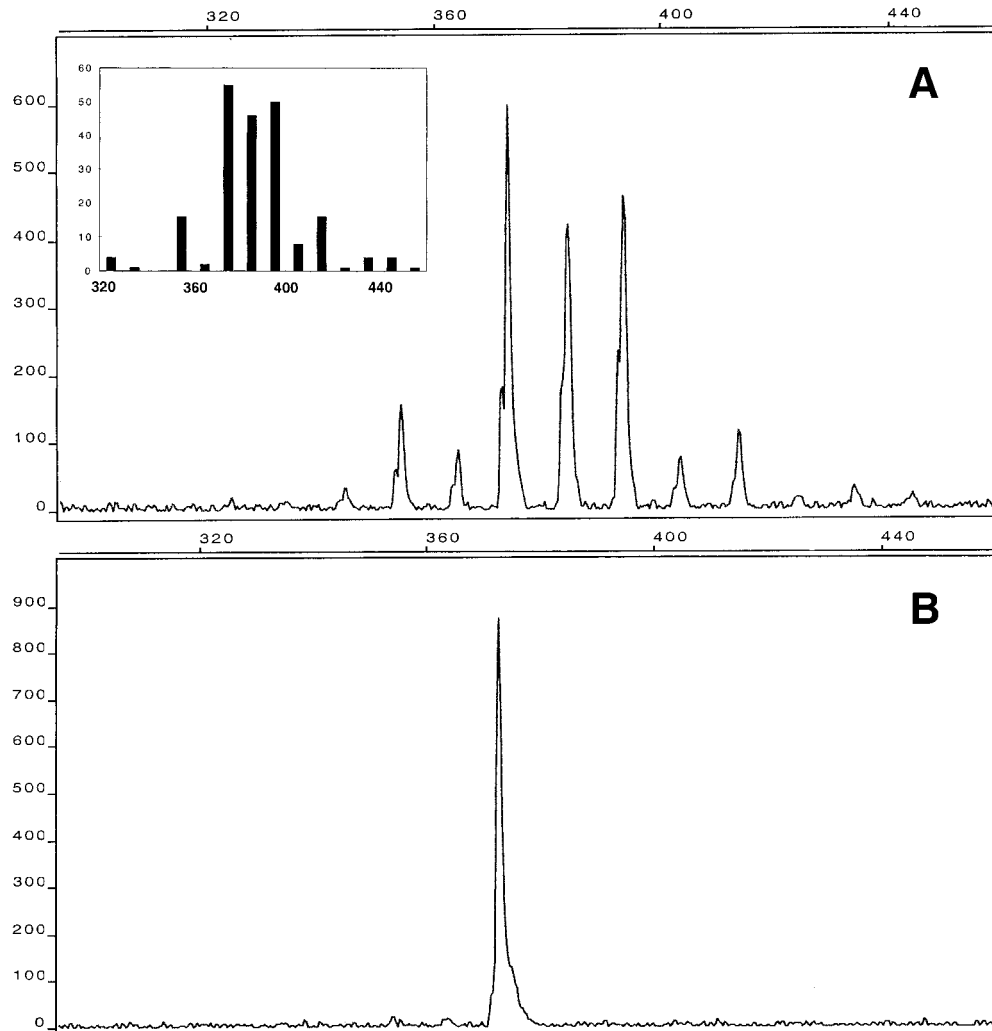


FIG. 2.—A, Fragment analysis of the tandem repeat region; PCR product from total DNA from the individual PD3. Lengths of fragments in base pairs and relative intensity of fluorescence are shown on the X and Y axes, respectively. Inset picture shows the distribution of the number of cloned PCR fragments of different lengths from PD3. B, Fragment analysis of the PCR product from one clone. Fragment length in base pairs and relative intensity of fluorescence are shown on the X and Y axes, respectively.

Ind.	A.t.	N	Repeat type sequence	Ind.	A.t.	N	Repeat type sequence
PD1 (D6)	i2	1	1111111110010000111010101010101	PD2 (D6)	46a	1	1111111100100000111010101011111010101010101
	8a	1	11111111100100000111010101010101		2a	1	111111110010000011101010101010101010101
	49a	1	111111111001000011101010101010101		51c	2	11111100100000111010101010101010101
	i4	5	111111110010000011101010101010101		i1	3	11111111111001000011101010101010101
	29a	1	11111001000000011101010101010101		5a	1	11111111110010000011101010101010101
	i5	1	1111111001000011101010101010101		19b	1	1111111111001000011101010101010101
	i6	1	1111110010000011101010101010101		i2	4	1111111111001000011101010101010101
	13a	1	1111100100000011101010101010101		i3	4	1111111110010000011101010101010101
	i10	1	1111111001000011101010101010101		39b	1	11111110010000011101010101010101
	i9	2	11111100100001110101010101010101		72b	1	111110010000111010101010101010101
	i8	4	11111001000001110101010101010101		79b	2	111111110010000011101010101010101
	40a	1	11110010000111010101010101010101		102b	1	111111100100000011101010101010101
	46a	1	1111111000001110101010101010101		25b	1	111111100100000011101010101010101
	i12	1	11111100100001110101010101010101		30a	1	1111001000011101010100010101010101
	i11	1	1111100100001110101010101010101		71b	1	11111111001000011101010101010101
	i16	3	1111100100001110101010101010101		1b	1	111111100101000011101010101010101
	i15	5	11110010000111010101010101010101		28c	1	111111100100001011101010101010101
	i17	2	11100100000111010101010101010101		48c	1	111111100100001110101010101010101
	23a	1	1110010000111010101010101010101		i4	33	11111110010000011101010101010101
	i18	1	1110010000111010101010101010101		30b	1	11111001000000011101010101010101
	42a	1	1100100000111010101010101010101		29b	2	111001000000001110101010101010101
	27a	2	11111100100001010101010101010101		22a	2	11111111001000011101010101010101
	i20	1	1110010000111010101010101010101		i7	1	111111100100001110101010101010101
	i19	3	1110010000111010101010101010101		i5	9	111111100100001110101010101010101
	41a	1	11000001110101010101010101010101		i6	1	111111100100001110101010101010101
PD3 (D6)	53c	1	1111100100000111010101010001101010101		81c	1	111110101000001110101010101010101
	58b	1	1111100100000111010101000111010101010101		54b	1	111110010000111010101010101010101
	52c	1	11111001000011101010100000111010101010101		4c	1	1111111100100001110101010101010101
	88c	1	111111111111001000001110101010101010101		i10	2	11111111001000001110101010101010101
	14c	1	1111111111001000001101110101010101010101		36c	1	1111110010000111010101010101010101
	23b	1	111111111100100000111010101010101010101		i9	9	1111110010000111010101010101010101
	i1	4	11111111110010000111010101010101010101		20b	1	111110010000111010101010101010101
	i3	3	111111111100100000111010101010101010101		17b	1	1111100100000111010101010101010101
	26c	1	11111100100001110101010101010101010101		i8	5	1111100100000111010101010101010101
	11a	1	11111111110010000111010101010101010101		64c	1	1111100100000111010101010101010101
	1c	1	11111111110010000111010101010101010101		58b	2	1111000100000111010101010101010101
	15c	1	111111111100100000111010101010101010101		24b	1	1111111111001000010101010101010101
	41a	1	11111110010000011101010101010101010101		99b	1	1111111100101101010101010101010101
	i4	12	11111110010000011101010101010101010101		i12	3	1111110010000111010101010101010101
	43b	1	11110010000111010101010101010101010101		i13	2	11111001000011101010101010101010101
	i7	1	111111100100001110101010101010101010101		i11	11	11111001000011101010101010101010101
	i5	5	111111100100001110101010101010101010101		52b	1	11110010000111101010101010101010101
	36b	1	111110010100001110101010101010101010101		86b	1	1111001000011101010101010101010101
	72b	1	111110010000111010101010101010101010101		28a	1	1111001000011101010100010101010101
	i10	1	111111100100001110101010101010101010101		i14	2	11110010000011101010101010101010101
	21a	1	111111100100001010101010101010101010101		27b	1	1100100000011101010101010101010101
	i9	24	11111100100001110101010101010101010101		38c	1	111111001000011101010101010101010101
	3a	1	11111001000011100101010101010101010101		44c	1	11111001001110101010101010101010101
	61b	1	11111001000011110101010101010101010101		18b	1	11111001000011101111111111111101
	i8	17	111110010000111010101010101010101010101		i16	24	11111001000011101010101010101010101
	36c	1	11111000101111010101010101010101010101		88b	1	111110000101110101010101010101010101
	10a	1	11111000100001110101010101010101010101		82b	1	111100100001110101110101010101010101
	7a	2	11110100100001110101010101010101010101		16a	1	111100100001110101010101010101010101
	79b	1	1101100100000111010101010101010101010101		i15	22	111100100001110101010101010101010101
	22c	1	111110011100100000111010101010101010101		9c	1	1111001000001110101010101010101010101
	43a	1	111110011000111010101010101010101010101		13b	1	1110100100001110101010101010101010101
	28b	1	11111001000011110101010101010101010101		47b	1	111001000011101001010101010101010101
	i13	1	111110010000111010101010101010101010101		i17	3	11100100000111010101010101010101010101
	i11	36	111110010000111010101010101010101010101		22c	1	11100100000111010101010101010101010101
	18b	1	111110010000101110101010101010101010101		81b	1	10110010000111010101010101010101010101
	57b	2	1111100010000111010101010101010101010101		11a	2	11110010000111010101010101010101010101
	i14	1	111100100000111010101010101010101010101		72c	1	11101000000111010101010101010101010101
	20a	2	111100010000111010101010101010101010101		25a	2	11001000000111010101010101010101010101
	3c	1	11111100100001110101010101010101010101		7a	1	11111001000001110101010101010101010101
	i16	12	111110010000111010101010101010101010101		101b	1	111100010000111010101010101010101010101
	i15	39	1111001000011101010101010101010101010101		57b	1	11100110000111010101010101010101010101
	66b	1	111100100001110101000011101010101010101		i20	4	111001000011101010101010101010101010101
	63b	1	111100100001110000101010101010101010101		i19	1	111001000011101010101010101010101010101
	i17	1	1110010000011101010101010101010101010101		33c	1	11100010000111010101010101010101010101
	78a	1	111110011000111010101010101010101010101		75c	1	11001000011101010101010101010101010101
	i18	1	111001000011101010101010101010101010101		14b	1	11001000000111010101010101010101010101
	5b	1	1111110010000111010101010101010101010101		7c	2	10010000011101010101010101010101010101
	49c	1	111100001110101010101010101010101010101		1a	2	1111100100000111010101010101010101010101
	53a	1	1110100001110101010101010101010101010101		4b	1	1111111001000001110101010101010101010101
	96a	1	111001100011101010101010101010101010101		20a	1	111110010000111010101010101010101010101
	i20	8	111001000011101010101010101010101010101		68c	1	1100100000011101010101010101010101010101
	i19	3	111001000011101010101010101010101010101		25c	1	1111111001000001110101010101010101010101
	48b	1	110010000011101010101010101010101010101		i21	1	111111101010101010101010101010101010101
	76c	1	1111100111010101010101010101010101010101		75b	1	1111111001000000110101010101010101010101
	84c	1	1111100100000101010101010101010101010101		18a	1	111111101010101010101010101010101010101
	24a	3	1111100111010101010101010101010101010101		21b	1	1110010000011101010101010101010101010101
	i21	1	111111101010101010101010101010101010101		45b	1	11101010101010101010101010101010101010101
	94c	1	1000011101010101010101010101010101010101		13a	1	11110010101010101010101010101010101010101
	60c	1	11101010101010101010101010101010101010101		88c	1	11101010101010101010101010101010101010101
	33a	1	1101010101010101010101010101010101010101		42b	1	11100100001110101010101010101010101010101
	44c	1	1111001001				

FIG. 3.—Array types found in the different individuals. Individual (sequence variant), array type name, number of clones, and repeat type sequence are given. Array types that are found in more than one individual in a pedigree have the prefix “i” in the array type name.

Ind.	A.t.	N	Repeat type sequence	Ind.	A.t.	N	Repeat type sequence	
Wolf1 (D6/W6)	59	1	1220102120000021212120110001100000001	Wolf2b (D6/W6)	35	1	12212000121202011000110000000000001	
	47	1	122120001212121202011000110100000001		32	1	12022120001212020110001100000000001	
	4	1	12200121202011000110000001100000001		42	1	122220012121202201100011000000001	
	i1w	1	1221200012121212020110001100000001		i4w	1	1222200121212020110001100000001	
	1	1	1221100012121212020110001100000001		45	1	1222012121212020110001100000001	
	6	1	1222120012120201100011000000000001		i3w	1	1220012121212020110001100000001	
	5	1	12212000121212120200001100000001		i2w	1	1220010121212020110001100000001	
	29	1	12122120001212020110001100000001		15	1	12222001212020110001100000001	
	2	1	12212001212120201100001000000001		31	1	12200121213020110001100000001	
	i3w	1	1220012121212020110001100000001		i6w	14	12200121212020110001100000001	
	13	1	1220012121202011010001100000001		26	1	12200121112020110001100000001	
	i2w	3	1220010121212020110001100000001		i7w	1	12100121212020110001100000001	
	63	1	212000121202011000110000000001		16	1	122002121202011000110000001	
	14	1	122120001212121202001100000001		6	1	1220012121202011000110000001	
	58	1	122001212121220111000110000001		2	1	1220012121202011000110000001	
	10	1	122001212120201100010000000001		34	1	1220010121212120201100000001	
	15	1	112200121212020110001100000001		10	1	122001213120201001100000001	
	23	1	12201220012121202011001100001		9	1	122001212120201100011100001	
	i6w	31	12200121212020110001100000001		36	1	122001212120201100011000001	
	18	2	12200101212120110001100000001		41	1	12200121212020110001100001	
	68	1	1222212001212120201100000001		7	1	12200121212020110001100001	
	27	1	1220012121202022120001212021		47	1	1220012121202011000000001	
	31	1	1220012121202011001200000001		37	1	21212020110001100000001	
	39	1	1220012121201110001100000001		17	1	12222001	
	43	1	1121212020110001100000000001		Wolf2c (D6/W6)	5	1	122120020012121202011000110100000001
	25	1	1221212120201100011000000001			47	1	12222120012121202011000110100000001
	61	1	1202212020110001100000001			18	1	12200101012121202011000110000000001
i9w	1	1220201100011000000001	17	1		122120012121202011000110100000001		
7	1	1220110100000001	i2w	1		1220010121212020110001100000001		
Wolf2a (D6/W6)	15	1	1220012121202011001202011000110000001	12		1	1121212121212020110001100000001	
	i1w	1	1221200012121212020110001100000001	3		1	122001212120201110001100000001	
	30	1	122120001212121202011001100000001	45	1	122001212120201100001100000001		
	88	1	122001212121212020110001100000001	4	1	22000121212020110001100000001		
	19	1	12202000121212020110001100000001	33	1	12202121212020110001100000001		
	i4w	5	1222200121212020110001100000001	i6w	9	12200121212020110001100000001		
	71	1	1222001212120201100011000000001	21	1	12200111212020110001100000001		
	35	1	1220012121212020110001000000001	16	1	1122221200012121200011000000001		
	43	1	1220011221202011000110000000001	48	1	11212121202011000110100000001		
	38	1	1221001212120201100011000000001	1	1	2200012121202011000110000001		
	4	1	1220012121212020110001100000001	23	1	122012121202011012110000001		
	i5w	1	1220001212120201100011000000001	35	1	1121212121202011000110000001		
	59	1	1220001202120201100011000000001	46	1	122001212120201100100000001		
	2	1	122200121212020110001100000000	10	1	122001212120110001100000001		
	64	2	12200121212020110011100000001	27	2	122001212020110001100000001		
	i6w	45	12200121212020110001100000001	42	1	112121212120201100011000001		
	57	1	12200121212010110001100000001	31	1	1220212020110001100000001		
	36	1	12200121212001110001100000001	i8w	1	12200121212020110000001		
	i7w	1	12100121212020110001100000001	44	1	11212001212020110001		
	31	1	11200121212020110001100000001	24	1	12212020001		
	47	1	122000121212020110001100001	Wolf3 (D6/W6)	i4w	1	1222200121212020110001100000001	
	69	1	122121212020110011000000001		43	1	1220200121212020110001100000001	
	92	1	122001212020110011000000001		i5w	2	122000121212020110001100000001	
	63	1	122001212120210000000001		30	1	120000121212020110001100000001	
	i8w	1	122001212120201100000001		i6w	11	12200121212020110001100000001	
	i9w	1	122020110001100000001		44	1	122011121202011000110000001	
	8	1	12200000001		22	1	100000002	
			38		1	12200012		

FIG. 3 (Continued)

Ny38 (Norrbottnspets, D8); Ny39 (Irish setter, D9); Ny41 (Siberian husky, D18); Ny66 (German shepherd, D10); Ny71 (German shepherd, D5); Sch3 (Schipperke, D6); Wolf1, Wolf2a, Wolf2b, Wolf2c, and Wolf3 (maternal pedigree, wolf of Russian origin, W6/D6). Blood was sampled for H54, PD1, PD2, PD3, Sch3, and Wolf3, muscle was sampled for Wolf1, liver was sampled for Wolf2a, Wolf2b, and Wolf2c, and hair was sampled for H9, H83, Ny34, Ny36, Ny38, Ny39, Ny41, Ny66, and Ny71.

#### DNA Preparation

DNA was obtained from blood samples using the Chelex procedure (Walsh, Metzger, and Higuchi 1991). DNA from muscle and liver samples was obtained using the following method: 1-cm<sup>3</sup> pieces were cut into slices, washed in 1 × SSC, and put in 400 µl of 150 mM NaAc (pH 7.0), 1.25 mg/ml proteinase K, 50 mM DTT, and

2% NP40 detergent. The samples were incubated at 37°C overnight and extracted twice with phenol/chloroform, and DNA was recovered by two rounds of ethanol precipitation. DNA was obtained from hairs using the following method: hairs were pulled from the individuals so that the bulbs were recovered and placed in 200 µl of 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, Calif.). The mixture was incubated at 56°C overnight and at 96°C for 10 min and was finally subjected to vortex mixing and centrifugation. The supernatant was used directly in the PCR amplification.

#### Direct DNA Amplification

The RS3 tandem-repeat 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



Ind.	A.t.	N	Repeat type sequence	Ind.	A.t.	N	Repeat type sequence
Ny34 (D1)	2	1	101010101110000101001001011010001	Sch3 (D6)	7b	1	101011010100010100100101001010101001
	1	2	10101111000010100100101011010001	3b	1	1010010100010000101001011010110101001	
Ny39 (D9)	1	1	110011001111000110001011011011001	18	1	11010100101010010010100101101011001	
	2	1	1000110011000011000101011011001	8	1	1010100101001001010010110110101001	
	5	1	11001110000110001011011011001	17	2	1010110101001010010010010110101101	
Ny41 (D18)	5	1	111010001010111000010100101001101	3	1	1010101001010010010100101101011001	
	9	1	100100011111000011010010100011101	24	1	10101001010010010100101101011001	
	10	1	11000111111000000110100101001101	4	1	10101001010010010100101101001001	
	5b	1	1100001111100000010100101001101	19	1	10010100100100001010010110101101	
	6	2	1100011111000000110100101001101	21	1	0100100110100100100101010101101	
	2	1	1001000111110000111010010101101	13	1	0100100101001101010010110101101	
H9 (D5)	24	3	111100101100110010101000001001011	7	1	101011010100101001001001011001	
	1	1	111011100110110101010100001001011	1	1	101001010010010100101101011001	
	5	2	1111001011001001010100001001011	8b	1	101001010010000101001010110101	
	8	1	1110011011010101010100001001011	23	1	10010101001000010101010101101	
	15	1	111101011001001010100001001011	15	1	100101001000010100101101011001	
	16	2	111100110110101010100001001011	10	8	10101001010010010010110101101	
	39	2	111011011010101010000101001011	14	2	10010100100001010010110101101	
	11	1	111001101101010101000010001011	6b	1	1010010100100001010110101101	
	10	16	11100110110101010100001001011	12	1	1000010100100001010010111101	
	18	2	11001101101010101000010001011	9	1	1010010100100100101010110101	
H83 (D5)	31	1	111111001010101010101010001001011	5b	1	00101001001010010110101101	
	32	1	11101001000010101110101010001001011	4b	1	111010111101	
	40	2	111010010000101010101010001001011	Ny66 (D10)	2	1	1000100010101010000001101000101
	36	2	1110010001010101010101010001001011	4	1	1010101010100000011010000101	
	43	2	1110010001010101010101010001001011	1	2	100010101010000001101000101	
	28	1	111001000101010101010100010001011	H54 (D8)	21	1	1110000000111000000001100111000011001
	37	12	11100100010101010101010001001011	10b	1	11111110100000000110011100011001	
	17	1	1110010001010101010101000001001010	6b	1	11101111010000000110011100011001	
	42	1	1110100100010101010101010001011	11b	1	1111111010000000110011100011001	
	9	1	111010001010101010101010001001011	3b	1	111010100000000110011100011001	
	47	1	10000000010100100000000001011	18	1	1110111011000001110011100011001	
	24	1	1100010101010101010001001011	4b	1	1110101011000001110011100011001	
	27	1	11100100010101001011	8b	1	111110100000000110011100011001	
Ny71 (D5)	16	9	110101110100010101010000010001011	14	1	111101011000001110011100011001	
	3	1	010111010100010101010000010001011	1b	1	11101011000001110011100011001	
	31	1	10101110100010101010000010001011	1	1	11101011000001110011100011001	
	30	1	01101110100010101010000010001011	22	1	11101011000001110000000001101	
	19	1	01001110100010101010000010001011	17	1	1110100000001100111000011001	
	20	1	1101011101000101010000010001011	9	1	111011000001110011100011001	
	38	4	0101110100010101010000010001011	9b	1	111010000000110011100111001	
	44	1	010111010000101010000010001011	7b	1	111000000001100111000011001	
	1	1	11110100010101010000010001011	5b	1	110000000001100111000011001	
	13	3	01011101000101010000010001011	2b	1	11010000000011001110011001	
	15	1	01011101000010101000010001011	Ny38 (D8)	2	1	1110100010001001001000000100001
	4	1	1110100010101010000010001011	4	1	111010010010001001000000100001	
	37	1	010100010101010000010001011	3	1	11101010001001001000000100001	
	36	1	1101010101010000010001011	1	1	111010001001001000000100001	
Ny36 (D7)	5	1	101010100111010101011000000001				
	2	1	101010100111010101011000000001				
	3	1	1010101010110001110000000001				

FIG. 3 (Continued)

TGA CAT GAG TTT ACG-3'). For control experiments and nested PCR, 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 analysis, WD5 was fluorescently labeled with 6-FAM dye label (PE Applied Biosystems, Foster City, Calif.). The PCR mixture consisted of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 2 µg/ml BSA, 0.2 mM of each dNTP, 0.1 µM of each primer, and 2 U of *AmpliTaq* DNA polymerase (Perkin Elmer, Norwalk, Conn.) in a volume of 50 µl. The amplification program consisted of predenaturation (94°C for 2 min) followed by 40 cycles of denaturation (94°C for 15 s), primer annealing (62°C for 30 s) and extension (72°C for 1 min), and a final extension (72°C for 10 min).

#### DNA Cloning and DNA Amplification of Cloned DNA

Amplified DNA was ligated into the pGEM-T vector (Promega, Madison, Wis.), transformed into *Escherichia coli* competent cells (Stratagene, La Jolla, Calif.), and spread according to the manufacturers' directions.

Individual bacterial clones were picked and taken directly to PCR amplification. 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 and amplification program were identical to those used for direct PCR, except that 35 cycles were used and primer annealing was performed at 69°C.

#### DNA Fragment Analysis and Sequence Analysis

For fragment analysis, 0.2 µl of amplification product was mixed with deionized formamide and fluorescently labeled size standard (Tamra 500, PE Applied Biosystems). For sequence analysis, 1 µl of amplification product was mixed with 3.2 pmol 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), except that primer annealing was performed at 55°C. The samples were analyzed on an ABI PRISM 377XL (PE Applied Biosystems) using 4% denaturing polyacrylamide gel according to the manufacturer's di-

**Table 1**  
**A Three-Generation Maternal Pedigree of Dogs Showing Array Types Found in More than One Individual**

INHERITED ARRAY TYPE	LENGTH (NO. OF REPEATS)	PD1		PD2		PD3	
		No. of Clones	Frequency (%)	No. of Clones	Frequency (%)	No. of Clones	Frequency (%)
i1 .....	35	—	—	3	1.4	4	1.8
i2 .....	34	1	<3	4	1.8	—	—
i3 .....	34	—	—	4	1.8	3	1.4
i4 .....	32	5	11.6	33	15.1	12	5.5
i5 .....	31	1	<3	9	4.1	5	2.3
i6 .....	31	1	<3	1	<1	—	—
i7 .....	31	—	—	1	<1	1	<1
i8 .....	30	4	9.3	5	2.3	17	7.8
i9 .....	30	2	4.6	9	4.1	24	11.0
i10 .....	30	1	<3	2	<1	1	<1
i11 .....	29	1	<3	11	5.0	36	16.5
i12 .....	29	1	<3	3	1.4	—	—
i13 .....	29	—	—	2	<1	1	<1
i14 .....	29	—	—	2	<1	1	<1
i15 .....	28	5	11.6	22	10.1	39	17.9
i16 .....	28	3	7.0	24	11.0	12	5.5
i17 .....	28	2	4.6	3	1.4	1	<1
i18 .....	27	1	<3	—	—	1	<1
i19 .....	26	3	7.0	1	<1	3	1.4
i20 .....	26	1	<3	4	1.8	8	3.7
i21 .....	25	—	—	1	<1	1	<1
Unique.....		11		74		48	
Total.....		43		218		218	

NOTE.—The number and relative frequency of clones containing each array type are shown. Also shown are the number of clones containing unique array types and the total number of analyzed clones.

rections. Genescan software (PE Applied Biosystems) was used for size calling and quantification of the DNA fragments, and 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 trinary code and aligning the codes using in-house-developed software. For pairwise sequence comparisons in the mutation analysis, computer programs tailored for comparing two array types were used. The program assumes that the

first array type is ancestral to the second and that the evolutionary events are point mutations, deletions, or insertions where the insert is adjacent to its template. The program gives an upper bound for the number of evolutionary events needed to explain the evolution of the first sequence into the second. The software tools for sequence comparisons on a Unix system are available on request.

#### PCR and Cloning Artifacts

In order to ascertain that the heteroplasmic variation found in the analyzed individuals was not caused by PCR or cloning, several experiments were carried

**Table 2**  
**A Three-Generation Maternal Pedigree of Wolves Showing Array Types Found in More than One Individual**

INHERITED ARRAY TYPE	LENGTH (NO. OF REPEATS)	WOLF1		WOLF2A		WOLF2B		WOLF2C		WOLF3	
		No. of Clones	Frequency (%)	No. of Clones	Frequency (%)	No. of Clones	Frequency (%)	No. of Clones	Frequency (%)	No. of Clones	Frequency (%)
i1w .....	34	1		1		—		—		—	
i2w .....	31	3		—		1		1		—	
i3w .....	31	1		—		1		—		—	
i4w .....	31	—		5		1		—		1	
i5w .....	30	—		1		—		—		2	
i6w .....	29	31	50	45	59	14	38	9	26	11	58
i7w .....	29	—		1		1		—		—	
i8w .....	23	—		1		—		1		—	
i9w .....	21	1		1		—		—		—	
Unique .....		25		21		19		23		5	
Total .....		62		76		37		34		19	

NOTE.—Wolf2a, Wolf2b, and Wolf2c are daughters of Wolf1, Wolf3 is the daughter of Wolf2a. The number and the relative frequency of clones containing each array type are shown. Also shown are the number of clones containing unique array types and the total number of analyzed clones.

out. First, from one individual, multiple blood samples were taken, DNA was extracted, and PCR amplifications and fragment analysis were performed (fig. 2A). The different blood samples gave identical fragment analysis profiles (data not shown). Second, PCRs performed with alternative primers and with nested primers yielded the same profiles (data not shown). Third, when the PCR products were cloned and the clones were PCR amplified, only one peak per clone was revealed by fragment analysis (fig. 2B). Fourth, when the cloned fragments of different lengths were counted, the distribution of the number of fragments of different lengths agreed with the fragment analysis profiles (fig. 2A). These experiments show that the majority of the different length types found were actually present in the original DNA extract and that even though the existence of length artifacts created by PCR errors is not completely ruled out, they are rare. Furthermore, these experiments show that the frequency distribution of the length types was faithfully transferred through the cloning step.

In order to estimate the maximum frequency of any possible PCR length artifacts not detected in the previous experiments, dilutions down to a few molecules were made on the DNA extracts and PCR amplifications were performed. Thereby, only a few different array types were start molecules in the PCR. From an extract from PD3 containing approximately 5–10 mtDNA molecules, 46 clones were sequenced and 5 different array types were found. When compared with the array types found in the dog pedigree, three of the five array types were identical to 1 of the 21 inherited array types, and two were unique array types. Only the two unique array types are potential artifacts, as an artifact mutation resulting in an array type previously found in more than one individual is unlikely. Similar dilutions were performed on extracts from PD1 and PD2. From the three samples, a total of 129 clones were analyzed, and only four unique array types were found. This indicates that if length artifacts do occur, the maximum error rate is approximately 3.1 artifact array types per 100 analyzed clones.

#### Frequency of Point Mutation Artifacts

To investigate the frequency of point mutation artifacts, the PCR product from one clone from PD3 was cloned once more in a fashion identical to that used before. Thirty-three clones were analyzed by PCR and sequence analysis. Six point mutations, two of which were in the informative position, were found in the 9,570 bp analyzed, yielding a point mutation artifact frequency of  $6.3 \times 10^{-4}$ . Assuming that the artifact frequencies are equal among sites, a frequency of 1.7 point mutation artifacts in the informative position per 100 clones is obtained.

## Results

### Intra- and Interindividual Fragment Length and DNA Sequence Variation

The RS3 tandem-repeat region was studied by fragment length analysis and DNA sequence analysis for 14

dogs and 5 wolves. All individuals showed length heteroplasmy caused by different numbers of 10-bp repeats in the tandem-repeat arrays (fig. 2A). All individuals analyzed, including maternally related ones, had different fragment analysis profiles (data not shown). A second level of variation was caused by the fact that the repeated 10-bp sequences were not perfect repeats. The dogs had two different 10-bp repeat types differing in one position, indicated in bold type: ACACGT**G**CGT and ACACGT**A**CGT. In the wolf pedigree, there was a second variable position, and a third repeat type, ACA-CAT**A**CGT, was present in these individuals. The three repeat types were designated 0, 1, and 2, respectively. The repeat types were distributed in different patterns, resulting in a secondary DNA sequence (fig. 3). Only the informative positions, i.e., those defining the two (for dogs) or three (for wolves) repeat types, were included in the analyses. Point mutations found in other positions were not taken into account. Within the individuals, several different arrays having different repeat type sequences (array types) were found, differing from each other by potential insertions or deletions (indels) of one or several repeats or point mutations in the informative position. The smallest array type found had 8 repeats, and the largest had 46 repeats. However, a very narrow distribution of the number of repeats was found in all individuals; 92% of the molecules had 25–35 repeats, and 97% had 20–40 repeats. Of 869 molecules studied, only 3 showed insertions or deletions of truncated repeats. These molecules were omitted from the analyses. It could be noted that in the wolves, there were two informative positions but only three repeat types instead of the possible four. This shows that positions 5 and 7 in the repeats are kept together as one unit in the length mutations.

All individuals from which several clones were analyzed had large numbers of different array types. Two individuals from the maternal pedigree of dogs were more thoroughly studied. Two hundred eighteen clones were analyzed for PD2 and PD3, and 85 and 61 array types were found in the two individuals, respectively (fig. 3). A sample coverage analysis (Chao and Lee 1992) indicated total numbers of 384 and 334 array types in the two individuals, respectively.

The distribution of the frequencies of different array types within the individuals varied considerably. Some individuals had one array type at a frequency of over 50% and all other array types at much lower frequencies, while other individuals had several array types at intermediate and more equal frequencies (fig. 3). For example, in three of the members of the wolf pedigree and in H9, a single array type was predominant and constituted 50%–60% of the molecules, while the other array types were present in frequencies under 10%. In PD2, PD3, and Ny71, on the other hand, there were three array types present in frequencies between 10% and 20%, and in H54, all 18 molecules analyzed were unique, indicating that all array types had frequencies under 15%.

A pairwise repeat type sequence comparison performed on all possible pairs of array types among the

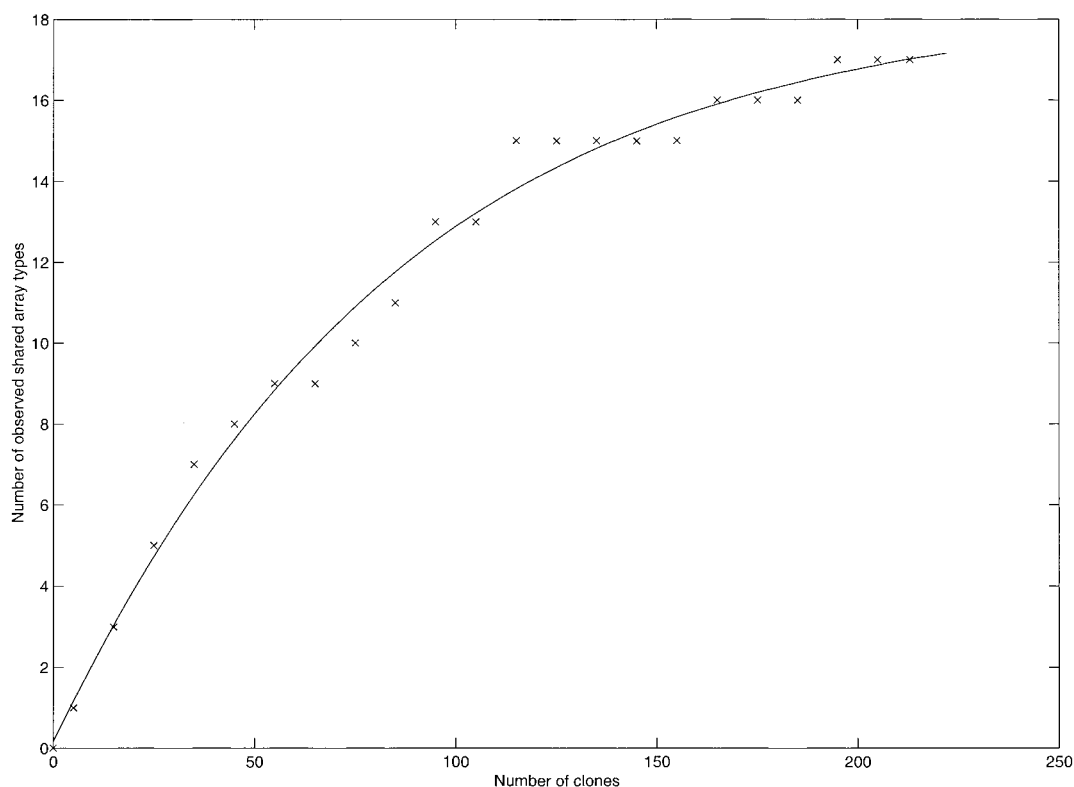


FIG. 4.—The number of array types found in both PD2 and PD3, plotted against the number of analyzed clones. The increase in the number of shared variants is shown for every 10 clones analyzed in PD2 and PD3.

unrelated individuals showed no array type matches, not even between individuals sharing the same nonrepetitive CR sequence variant. Three German shepherds sharing the same sequence variant (D5) were also divergent. It can therefore be concluded that the interindividual variation of the sequence of repeat types is larger than that of the nonrepetitive CR sequence. Within individuals, a common pattern in the repeat type sequence was found in the different array types, and usually each array type differed by a single indel or point mutation from one or several other array types. In most cases, more similarity between the array types was found in the 3' end than in the 5' end.

Very different structures in the patterns of the repeat type sequences were found when some individuals were compared. For example, in molecules found in PD1, PD2, and PD3, the 5' end consisted mostly of a consecutive row of repeats of type 1, and the 3' end had a pattern of alternating type 0 and type 1 repeats, while among the wolves, the 5' end consisted mostly of alternating repeats of types 1 and 2 (GenBank accession number AF202894), and the 3' end consisted of a consecutive row of repeats of type 0. In Sch3, an alternating pattern of repeat types 0 and 1 was found all through the arrays.

#### Inheritance of Array Types in the Maternal Pedigrees and Identification of Mutated Array Types

For the dog pedigree, 43, 218, and 218 clones, respectively, were sequenced from the three animals PD1,

PD2, and PD3 (fig. 3 and table 1). In all individuals, a few array types were more common, with frequencies of up to 18%, while the majority of the array types were found only once, indicating that the frequencies of most array types were lower than 1%. The most frequent array types were inherited from mother to offspring, but the frequencies varied from generation to generation. The 11 array types that were common to PD1, PD2, and PD3 constituted 65%, 56%, and 73% of the molecules found in these individuals, respectively. The majority of the array types were found in only one individual. For example, 43 of the 61 array types found in PD3 were unique. The array types that were unique to one individual constituted 19%–33% of the molecules in the three individuals. In the wolf pedigree, only one array type was common to all individuals (table 2). This type constituted 26%–59% of the molecules found in the individuals. All other array types were found in only one to three individuals, and at low frequencies. Also in this pedigree, the proportion of the inherited array type fluctuated considerably between generations. This fluctuation indicates a narrow genetic bottleneck in the germ line. The number of segregating units in the germ line in mice has been calculated to be about 200 molecules using the variance in the frequency of mtDNA types in the offspring of female mice, and similar values were found for human pedigrees (Jenuth et al. 1996). When the same analysis was performed on Wolf1 and its offspring Wolf2a, Wolf2b, and Wolf2c (the dog pedigree could not be analyzed, as there was only one offspring



[illegible]

## B

Uni.	Inh.	N	Alignment	Events	e
78a	i16	12	1111110010000111011010101101 111111001100011101101010-101	dp	2
58b	i11	36	1111110010-000111010101-----0101101 111111001000001110101010011101010101101	ii	2
66b	i15	39	11111001000011101-----01010101101 1111100100001110101000011101-----	id	2
61b	i11	36	1111110010-0001110101010101101 11111100100000111101010101101	ip	2
	i8	17	1111110010000011--1010101010101 11111100100000111101010101101--	id	2
18b	i16	12	11111100100001--11011010101101 1111110010000101110-1010101101	id	2
	i11	36	11111100100001110101010101101 11111100100001011101010101101	pp	2
21a	i16	12	11--11110010000111011010101101 11111110010000101011010101101	ip	2
	i5	5	11111110010000011101010101101----- 1111111001000-----01010110101101	id	2
	i20	8	1111-----0010000111011010101101 1111111001000010111010101101	ip	2
14c	i8	17	11111-----10010000011--1010101010101 11111111111100100000110111010101010101	ii	2
	i3	3	1-11111111110010000011--1010101010101 11111111111100100000110111010101010101	ii	2
	i4	12	111--111110010000011--1010101010101 11111111111100100000110111010101010101	ii	2
22c	i8	17	11111100-----100000111010101010101 11111100111001000000111010101-----1	id	2
	i8	17	11111-----10010000011--1010101010101 1111111111110010000010111010101010101	ii	2
	i3	3	1-11111111110010000011--1010101010101 11111111111100100000110111010101010101	ii	2
	i4	12	111--111110010000011--1010101010101 11111111111100100000110111010101010101	ii	2
33a	i1	4	11111111111100100001110101010101101 -----110101010101---	dd	2
	i11	36	11111100100001110101010101101 -----110101010101---	dd	2
	i19	3	111001000011101010101010101 -----110101010101---	dd	2
	i5	5	1111111001000001110101010101101 -----110101010101---	dd	2
	i8	17	111111001000001110101010110101 -----110101010101---	dd	2
	i3	3	111111111100100000111010101010101 -----110101010101---	dd	2
	i4	12	11111110010000011101010101010101 -----110101010101---	dd	2
	i15	39	1111100100001110101010101101 -----110101010101---	dd	2
	i9	24	111111100100001110101010101101 -----110101010101---	dd	2
52c	i15	17	11111001000011101010101-----101101 11111001000011101010100000111010101-10-101	idd	3
49c	i8	17	1111110010000011101010101--10101 11111-----000011101010101-010101	did	3
36c	i5	6	1111-----1111001000011101010101101 --1111110001011110-----1010101010101	did	3
53c	i16	12	11111100100100-----00-111011010101101 1111110010000011101-01010110100011101010101101	idd	4

FIG. 5 (Continued)

per generation), the size of the germ line bottleneck was calculated to be 59.6 molecules, indicating that the number of segregating units in dogs and wolves is similar to those found in mice and humans.

The large number of array types in PD2 and PD3 and the presence of a germ line genetic bottleneck in-

dicate that most of the unique array types were not inherited from PD2 to PD3 but were created by mutations. Furthermore, the fact that the frequencies of the most common array types, though fluctuating, remain at the same level over three generations in the pedigrees shows that the mutations are somatic, as a high frequency of germ line mutations would result in a decrease in the frequency of the common array types. In order to establish that practically all of the analyzed molecules which had been inherited from PD2 to PD3 had actually been identified and that the other molecules could be assumed to be somatically mutated, two analyses were performed. First, the number of array types shared between mother and offspring was plotted against the increasing number of analyzed clones (fig. 4). The experimental curve shows that a large proportion of the inherited array types had been found by the analysis of 218 clones in PD2 and PD3 and, importantly, that all of the most frequent inherited array types had been found. A very conservative extrapolation of the curve, under the assumption of a linear increase using the tangent of the last few clone pairs, shows that to find five more inherited array types, a total of 500 clones must be analyzed. A number of inherited array types considerably larger than the 17 found to be common for PD2 and PD3 is therefore not consistent with a germ line bottleneck size of about 200 molecules. Second, a sample coverage analysis (Chao et al. 2000) indicated that 35 array types were shared between PD2 and PD3, to be compared with the total number of 334 array types estimated to be present in PD3. In both analyses, an overestimate of the number of shared array types is probable, as array types identified as inherited but found only once in PD3 may have arisen independently by somatic mutations. The analyses show that only a minority of the array types in PD3 had been inherited from PD2 and that, among the array types in PD3 found in this study, practically all of those inherited from PD2 to PD3 had been identified. This implies that the other array types, which were found in PD3 but not in PD2, were mutated forms of the inherited array types. Therefore, it is indicated that as much as 20%–30% of the molecules in the individuals are somatically mutated.

#### Somatic Mutations: Test of Replication Slippage Mutation Model

In order to examine the mutation mechanism responsible for the length mutations, and to study the na-

FIG. 5.—A, The possible explanations for the 37 one-step mutations found in PD3. The unique array type (Uni.), the inherited array type (Inh.) which is the possible origin of the unique array type, the frequency of the inherited array type among 218 clones, the alignment showing the mutation, the type of mutation event (insertion, deletion, or point mutation), and the position for the mutation are given. The 5' end, the 3' end, and the middle of an array are defined by dividing the molecule into three equally long parts. When several positions are possible for the indel, the most central one is chosen. When several inherited array types can be the origin of a unique array type, the type and position of the mutation are given for the most common inherited array type only. Note that the inherited molecule is on the top and the mutated molecule is on the bottom in the pairwise alignments. B, The possible explanations for the nine two-step mutations, the three three-step mutations, and the single four-step mutation found in PD3. For the three- and four-step mutations, only one of several possible mutations are shown. The unique array type (Uni.), the inherited array type (Inh.) which is the possible origin of the unique array type, the frequency of the inherited array type among 218 clones, the alignment showing the mutations, the type of mutation event (insertion, deletion, or point mutation), and the number of mutation events (e) are shown. Note that the inherited molecule is on the top and the mutated molecule is on the bottom in the pairwise alignments.

ture of the mutations, a comprehensive study was performed on the mutated array types found in PD3. To be able to study a mutation, it is necessary to identify the original and the mutated molecule types. Here, this was possible, as it could be presumed that all the original array types, transferred through the germ line, had been found and that all other array types were the results of somatic mutations in the original ones. As it could not be known beforehand which of the inherited array types was the origin of which mutated array types, all inherited array types had to be compared with all mutated array types.

Two different types of mechanisms can be expected to be involved in the length mutations in the repeat arrays: replication slippage and recombination. In the case of recombination, the insertions and deletions can occur anywhere in the arrays, and the inserted elements can be expected to have the sequence of any part of the molecules found in the population of inherited molecules. In the case of replication slippage, the outcome of the mutation is more restricted, as the insertions are always situated directly adjacent to (in tandem with) their template sequences. As the replication slippage model is the most conservative of the models, and any result consistent with replication slippage can also be explained by recombination, only the replication slippage model was tested.

The test of the replication slippage model was performed by aligning the inherited and the mutated array types in pairs and for each pair investigating whether the existence of the mutated array type could be explained by one or more mutations (point mutations, deletions or insertions caused by replication slippage) in the inherited array type. The explanation requiring the smallest number of mutations was chosen for each mutated array type. It was presumed that all array types in PD3 that were also found in PD2 were inherited, except those that were found only once in PD3, as these could have been created by independent somatic mutations (table 1). All molecules having a unique array type and the inherited array types i7, i10, i13, i14, i17, i18, and i21 were thus regarded as mutated molecules, giving 50 array types regarded as mutated. As these array types presumably are somatic-mutation variants of the inherited array types, the frequency of somatic mutation was set at  $50/218 = 0.229$  (the number of mutated array types per 218 analyzed clones). At this frequency of mutation, there is a high probability that some molecules are mutated more than once. For example, under the assumption of a population of constant size where all molecules are equally likely to become mutated, the number of mutations in the different molecules would be Poisson distributed. However, in this case, the population has experienced an expansion from a few copies in the single germ cell, and a distribution skewed toward more molecules having several mutations may be expected.

When the mutated array types in PD3 were compared with the inherited array types, the existence of all 50 mutated array types could be explained by one to four slippage-induced indels or point mutations (fig. 5).

Thirty-seven of the mutated array types (43.7 expected according to the Poisson distribution model) could be explained by a single mutation, nine could be explained by two mutations (5.7 expected), three could be explained by three mutations (0.49 expected), and one could be explained by four mutations (0.03 expected). Thus, the results are in accordance with a model in which mutations are caused exclusively by replication slippage or point mutations.

The analysis was also performed on the array types in PD2. In this case, there were only 43 clones analyzed in the mother, PD1, and all array types inherited from PD1 to PD2 had probably not been identified. Instead, it was presumed that all array types that either were unique to PD2 or were found only once were mutated array types and that the remaining array types were inherited from PD1. This gave 16 inherited array types and 69 mutated types and a frequency of somatic mutation of  $69/218 = 31.7\%$ . When the presumably mutated array types were aligned and compared with the inherited ones, 47 array types could be explained by one mutation (56.8 expected according to the Poisson distribution model), 18 could be explained by two mutations (10.8 expected), 4 could be explained by three mutations (1.4 expected), and 0 had to be explained by four or more mutations (0.13 expected) (data not shown).

#### Mutation Characteristics

The possible explanations for the mutations in all unique array types in PD3 are shown in figure 5. In some cases, it is impossible to determine from which of two or more inherited array types a unique array type stems, as mutations in two different inherited array types may result in the same new type. This is especially the case with the array types that have been altered in the 5' end, with its row of up to 14 consecutive type 1 repeats, and for deletions. However, for 19 of the 37 array types that had been mutated once, there was only one most-parsimonious explanation. When the individual pairwise comparisons between the inherited and the mutated array types are studied in PD3, the insertion and deletion events can be studied in detail. The length and the repeat type sequence of the indel elements and the approximate position of the mutation event can be determined. The characteristics of the mutations found in the molecules mutated once in PD2 and PD3 are given in tables 3 and 4. Mutations were found in all parts of the arrays, but at a higher frequency in the 5' end. Almost half of the insertions and deletions were one or two repeats long. There was a small predominance of deletions over insertions and a 2.5-fold excess in the total number of deleted repeats compared with inserted repeats. The excess of deleted repeats could be attributed to 10 array types created by large deletions 9–17 repeats long. Fifteen point mutations were found in PD2 and PD3 in the informative position. Considering the expected PCR mutation frequency (7.4 point mutations per 436 clones, see *Materials and Methods*) and the fact that most of the point mutations were found in the 5' end, it is indicated that some of the point mutations are au-

**Table 3**  
**Summary of Mutation Characteristics**

INDIVIDUAL	TYPE OF MUTATION	NO. OF MUTATIONS	POSITION			NO. OF REPEATS
			5' End	Middle	3' End	
PD3 .....	Insertion	14	9	2	3	41
	Deletion	17	10	4	3	80
	Point mutation	6	3	2	1	—
PD2 .....	Insertion	18	6	9	3	47
	Deletion	20	7	7	6	143
	Point mutation	9	4	3	2	—
Wolf1, Wolf2a, and Wolf3 .....	Insertion	8	7	1	0	15
	Deletion	4	2	2	0	34
	Point mutation	4	2	2	0	—
H9 and H83 .....	Insertion	4	1	1	2	8
	Deletion	3	2	1	0	17
	Point mutation	0	0	0	0	—
All .....	Insertion	44	23	13	8	111
	Deletion	44	21	14	9	274
	Point mutation	19	9	7	3	—

NOTE.—The numbers of insertions, deletions, and point mutations, the numbers of mutations at different positions, and the total numbers of inserted and deleted repeats are given for the individuals studied. For the statistics obtained for the data, in the cases in which more than one explanation was possible for a mutation (for the type of mutation or for the number of repeats of the indel element), the explanation that involves the most common of the inherited molecules was chosen. The 5' end, the 3' end, and the middle of an array were defined by dividing the molecule into three equally long parts. When several positions were possible for the indel, the most central one was chosen.

thentic and that point mutations, as well as length mutations, are more frequent in the 5' end.

In order to elucidate whether the pattern of mutation found in PD2 and PD3 is general for the region or is influenced by the repeat type sequence of the molecules, e.g., by long stretches of identical repeats, the mutation analysis was performed on individuals having other array types: Wolf1, Wolf2a, and Wolf3 in the wolf pedigree, and H9 and H83, two unrelated German shepherds of sequence variant D5. These individuals had large proportions (44–59%) of one array type. It was presumed that this array type was inherited and that the array types that were unique to the individuals and were found only once were mutated molecules. The unique array types were aligned with and compared with the majority type, and the unique types that could be explained by one mutation in the majority type were iden-

tified (data not shown). The pattern of mutations in these individuals was similar to that found in PD2 and PD3 (tables 3 and 4).

#### Germ Line Mutations

The array types in PD3 that have been inherited from PD2 have been transferred in the germ line. The differences found between these array types are therefore the result of previous germ line mutations. An alignment of the 11 array types that were presumed to be inherited is shown in figure 6. The array types fall into three groups, each composed of array types that differ only in the 5' end. The differences between the array types can be explained exclusively by indels. Approximately the same pattern of mutations as in the study of the somatic mutations is found: most of the

**Table 4**  
**The Numbers of Insertions and Deletions of Different Sizes (Numbers of Repeats)**

INDIVIDUAL	TYPE OF MUTATION	NO. OF REPEATS								
		1	2	3	4	5	6	7	8	>8
PD3 .....	Insertion	4	4	0	3	1	2	0	0	0
	Deletion	4	4	0	2	2	1	1	1	2
PD2 .....	Insertion	7	3	5	1	1	0	0	0	1
	Deletion	4	2	1	1	1	1	2	0	8
Wolf1, Wolf2a, and Wolf3 .....	Insertion	2	5	0	1	0	0	0	0	0
	Deletion	0	1	0	0	0	1	0	1	1
H9 and H83 .....	Insertion	2	1	0	1	0	0	0	0	0
	Deletion	1	0	0	1	0	0	0	0	1
All .....	Insertion	15	13	5	6	2	2	0	0	1
	Deletion	9	7	1	4	3	3	3	2	12

NOTE.—For the statistics obtained for the data, in the cases in which more than one explanation was possible for a mutation (for the type of mutation or for the number of repeats of the indel element), the explanation that involved the most common of the inherited molecules was chosen.



Group	Repeat type	Frequency (%)	Repeat type sequence
1	i1	1.8	111111111110010000-11101010101011--01
	i5	2.3	----11111110010000-11101010101011--01
	i9	11.0	-----1111110010000-11101010101011--01
	i11	16.5	-----111110010000-11101010101011--01
	i15	17.9	-----11110010000-11101010101011--01
	i19	1.4	-----1110010000-11101010101011--01
2	i16	5.5	-----1111110010000-11101-10101011--01
	i20	3.7	-----11110010000-11101-10101011--01
3	i3	1.4	--1111111111001000001110101010--110101
	i4	5.5	----1111111001000001110101010--110101
	i8	7.8	-----11111001000001110101010--110101

FIG. 6.—An alignment of the 11 array types presumed to be inherited from PD2 to PD3. Group according to alignment, repeat type, frequency of array type among PD3 clones, and repeat type sequence are given.

mutations take place in the 5' end, and the mutations which can be easily identified, those that are not in the 5' end, are one or two repeats long.

## Discussion

A large amount of genetic variation was found in the region. When the array types were compared among the unrelated individuals, no array type matches were found. Furthermore, the frequency of the majority array type varied considerably among the individuals, and the array types had very different repeat type sequences in different individuals. Taken together, these observations indicate a high rate of germ line genetic evolution.

The proportion of molecules in PD3 with somatic mutations was calculated to be 21.2% (compensated for PCR point mutation artifacts), which is a very high figure compared with other estimates of the level of somatic mutation in mitotic tissue. In tandem repeat arrays in position RS5 in lagomorphs, the frequency of somatically mutated molecules was estimated to be between 1% and 5% (Casane et al. 1994). In the human brain, which is of a nonmitotic tissue type, 35%–45% of the mtDNA molecules had point mutations in the CR (Jazin et al. 1996). However, in blood, the frequency of point mutations could not be distinguished from the PCR artifact background.

The high proportion of somatically mutated molecules offered an opportunity to study the mechanisms causing the somatic mutations and their nature and position. The mutation analysis performed showed that all length mutations could be explained by replication slippage. Although recombination cannot be excluded by this study, direct evidence for recombination in vertebrate mtDNA has not been found. We therefore believe that, in light of the data presented here, replication slippage must be considered the most plausible explanation for the length polymorphism observed in these tandem arrays.

The study of the characteristics of the mutations showed a higher frequency of both length and point mutations in the 5' ends of the arrays than in the 3' ends. This observation agrees with the fact that, when compared within individuals, the array types were mostly more easily aligned in the 3' ends than in the 5' ends. The numbers of insertions and deletions were approximately the same, but the total number of deleted repeats was more than twice as large as the number of inserted

repeats. This difference could be attributed to 11 large deletions between 9 and 17 repeats long. The cause of this is unclear. If the mechanisms for mutations are identical in somatic cells and in germ line cells, this would result in a gradual loss of repeats in the arrays over generations. However, in a study of tandem repeats in the RS5 position in lagomorphs, it was shown that the arrays tend to lose repeats in somatic cells, but not in the gonads (Casane et al. 1997), and this could also be the case in dogs. The fact that the number of repeats in the arrays was narrowly distributed around 30 may be explained by an equilibrium between the rates of insertions and deletions in the germ line (Kruglyak et al. 1998). In this study, the mutations were studied for three groups of dogs or wolves having different repeat type sequences in their array types. The same general pattern of mutation was found in all three groups, which indicates that the mutation characteristics found are general for this region and not to a high degree dependent on the repeat type sequence of the arrays.

The position of the repeat locus, just at the end of the H-strand replication, could be one explanation for the high mutation rate. It is plausible that there is a better possibility for slippage as the replication nears completion. Furthermore, the RS3 array in pigs has been shown to be transcribed at the initiation of H-strand replication (Ghivizzani et al. 1993). DNA polymerase errors near the RNA-DNA transition may therefore be a cause of mutations. Furthermore, the arrays have self-complementary sequences capable of forming secondary structures, which could favor slippage by increasing the stability of the slipped strand or by blocking the polymerase. The fact that most of the mutations were found at one end of the array could be explained by the asymmetry in the replication, as only the H-strand replication originates in this region.

The differences among the array types identified as inherited from PD2 to PD3, which are caused by germ line mutations, could all be explained by insertions or deletions, and approximately the same types of length mutations as in the somatically mutated molecules were found. The higher frequency of germ line mutations in the 5' end is mirrored by the fact that when array types are compared among individuals (when alignments are possible), the array types are more similar in the 3' end than in the 5' end. The mitochondrial replication mechanisms are believed to be the same in somatic cells and



in germ cells, and it is probable that the findings obtained in the studies of the somatic mutations can be applied to the germ line mutations to a great extent. Therefore, replication slippage seems probable as the main source of length mutations in the germ line cells.

In the pedigrees, a considerable amount of fluctuation in the frequencies of the inherited array types was found between the generations. Furthermore, very different proportions of the majority array type were found among unrelated individuals. These facts indicate that in dogs as in other mammals, there is a relatively narrow bottleneck in the transfer of mtDNA molecules from mother to offspring in the female germ line, which causes genetic drift in the mtDNA population. This could explain the fact that the array types have very similar repeat type sequences within individuals, while the difference between individuals is sometimes large. If a single array type becomes predominant within an individual because of random drift, the majority of new array types will be mutated forms of this array type. Array types which do not originate from the predominant one will gradually be lost, and a relatively homogenous population of molecules is thus formed.

Mostly, it was not possible to align the array types from unrelated individuals without introducing a large number of indels (data not shown). Obviously, the evolution of the arrays is so rapid that comparisons of the array types is difficult even for individuals having the same nonrepetitive CR sequence variant. The individuals showing the greatest similarity were H9, H83, and Ny71, which are all German shepherds sharing the same sequence variant, D5. This similarity probably reflects a recent last common ancestor for these individuals, possibly in the last few hundred years. Preliminary analyses of genetic distance between individuals based on the mutation patterns found in this study showed that for individuals of different breeds, the array types were too dissimilar to give a useful result, but for the German shepherds, meaningful distance values were obtained (data not shown). This indicates that the analysis of these arrays may be used for phylogenetic studies on a timescale down to a few hundred years or less. However, more must be known about how rapid and homogenous among lineages the repeat type sequence evolution is before this possibility can be evaluated.

In conclusion, we propose that the following mechanisms are involved in the evolution of the tandem repeat arrays. Insertions and deletions are generated by replication slippage at a high rate, and point mutations occur at a lower rate, resulting in a heteroplasmic population of molecules. Most indels are one or two repeats long, but longer ones are not rare. The mutations are formed during the initiation or termination of the H-strand replication, and mutations are therefore most frequent in the 5' ends of the arrays. One reason for the high rate of length mutation is the capability of the arrays to form secondary structures, which may stabilize slipped strands or block the polymerase during replication. Because of a germ line bottleneck, the different array types are the subject of genetic drift, resulting in fluctuations in the proportions of the array types be-

tween generations. This leads to the extinction of some array types, while others become more frequent, resulting in homogenization of the population of molecules and fast evolution of the repeat type sequence.

## Acknowledgments

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). We would also like to thank Frode Lingaas for providing material from pedigree.

## LITERATURE CITED

- BROUGHTON, R. E., and T. E. DOWLING. 1997. Evolutionary dynamics of tandem repeats in the mitochondrial DNA control region of the minnow *Cyprinella spiloptera*. *Mol. Biol. Evol.* **14**:1187–1196.
- BUROKER, N. E., J. R. BROWN, T. A. GILBERT, P. J. O'HARA, A. T. BECKENBACH, W. K. THOMAS, and M. J. SMITH. 1990. Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. *Genetics* **124**:157–163.
- CASANE, D., N. DENNEBOUY, H. DE ROCHAMBEAU, J. C. MOUNOLOU, and M. MONNEROT. 1994. Genetic analysis of systematic mitochondrial heteroplasmy in rabbits. *Genetics* **138**:471–480.
- . 1997. Nonneutral evolution of tandem repeats in the mitochondrial DNA control region of lagomorphs. *Mol. Biol. Evol.* **14**:779–789.
- CHAO, A., and S.-M. LEE. 1992. Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* **87**:210–217.
- CHAO, A., W.-H. WANG, Y.-C. CHEN, and C.-Y. KUO. 2000. Estimating the number of shared species in two communities. *Statistica Sinica* (in press).
- CLAYTON, D. A. 1991. Nuclear gadgets in mitochondrial DNA replication and transcription. *Trends Biochem. Sci.* **16**:107–111.
- FUMAGALLI, L., P. TABERLET, L. FAVRE, and J. HAUSSER. 1996. Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. *Mol. Biol. Evol.* **13**:31–46.
- GHIVIZZANI, S. C., S. L. MACKAY, C. S. MADSEN, P. J. LAIPIS, and W. W. HAUSWIRTH. 1993. Transcribed heteroplasmic repeated sequences in the porcine mitochondrial DNA D-loop region. *J. Mol. Evol.* **37**:36–37.
- HOELZEL, A. R., J. M. HANCOCK, and G. A. DOVER. 1993. Generation of VNTRs and heteroplasmy by sequence turnover in the mitochondrial control region of two elephant seal species. *J. Mol. Evol.* **37**:190–197.
- HOELZEL, A. R., J. V. LOPEZ, G. A. DOVER, and S. J. O'BRIEN. 1994. Rapid evolution of a heteroplasmic repetitive sequence in the mitochondrial DNA control region of carnivores. *J. Mol. Evol.* **39**:191–199.
- JAZIN, E. E., L. CAVELIER, I. ERIKSSON, L. ORELAND, and U. GYLLENSTEN. 1996. Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **93**:12382–12387.
- JENUTH, J. P., A. C. PETERSON, K. FU, and E. A. SHOUBRIDGE. 1996. Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nat. Genet.* **14**:146–151.
- KRUGLYAK, S., R. T. DURRETT, M. D. SCHUG, and C. F. AQUADRO. 1998. Equilibrium distributions of microsatellite

- repeat length resulting from a balance between slippage events and point mutations. *Proc. Natl. Acad. Sci. USA* **95**: 10774–10778.
- LEVINSON, G., and G. A. GUTMAN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**:203–221.
- MIGNOTTE, E., M. GUERIDE, A. M. CHAMPAGNE, and J. C. MOUNOLOU. 1990. Direct repeats in the non-coding region of rabbit mitochondrial DNA. Involvement in the generation of intra- and inter-individual heterogeneity. *Eur. J. Biochem.* **194**:561–571.
- RAND, D. M., and R. G. HARRISON. 1989. Molecular population genetics of mtDNA size variation in crickets. *Genetics* **121**:551–569.
- SACCONE, C., G. PESOLE, and E. SBISA. 1991. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *J. Mol. Evol.* **33**:83–91.
- SAVOLAINEN, P., B. ROSEN, A. HOLMBERG, T. LEITNER, M. UHLEN, and J. LUNDEBERG. 1997. Sequence analysis of domestic dog mitochondrial DNA for forensic use. *J. Forensic Sci.* **42**:593–600.
- VILA, C., P. SAVOLAINEN, J. E. MALDONADO, I. R. AMORIM, J. E. RICE, R. L. HONEYCUTT, K. A. CRANDALL, J. LUNDEBERG, and R. K. WAYNE. 1997. Multiple and ancient origins of the domestic dog. *Science* **276**:1687–1689.
- WALBERG, M. W., and D. A. CLAYTON. 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. *Nucleic Acids Res.* **9**:5411–5421.
- WALSH, P. S., D. A. METZGER, and R. HIGUCHI. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**:506–513.
- WILKINSON, G. S., F. MAYER, G. KERTH, and B. PETRI. 1997. Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. *Genetics* **146**:1035–1048.

ROSS CROZIER, reviewing editor

Accepted November 11, 1999