J Forensic Sci, March 2009, Vol. 54, No. 2
doi: 10.1111/j.1556-4029.2008.00952.x
Available online at: www.blackwell-synergy.com

Kristen M. Webb,[1] *Ph.D. and Marc W. Allard,*[2] *Ph.D.*

Mitochondrial Genome DNA Analysis of the Domestic Dog: Identifying Informative SNPs Outside of the Control Region*

ABSTRACT: While the mitochondrial control region has proven successful for human forensic evaluations by indicating ethnic origin, domestic dogs (*Canis lupus familiaris*) of seemingly unrelated breeds often form large groups based on identical control region sequences. In an attempt to break up these large haplotype groups, we have analyzed the remaining *c.* 15,484 base pairs of the canine mitochondrial genome for 79 dogs and used phylogenetic and population genetic methods to search for additional variability in the form of single nucleotide polymorphisms (SNPs). We have identified 356 SNPs and 65 haplotypes in the remainder of the mitochondrial genome excluding the control region. The exclusion capacity was found to be 0.018. The mitochondrial control region was also evaluated for the same 79 dogs. The signals from the different fragments do not conflict, but instead support one another and provide a larger fragment of DNA that can be analyzed as forensic evidence.

KEYWORDS: forensic science, canine, mitochondrial genome, control region, SNP, haplotype, haplogroup

Hair, both human and animal, is often found as evidence in criminal investigations. Because hair is a composite of dead cells, the DNA contained in even fresh hair samples can be degraded (1). Each cell contains only two copies of the nuclear genome, but a second genome is also present in much higher copy numbers, the mitochondrial genome (mtGenome). Mitochondria are organelles responsible for many metabolic tasks within and between cells. There are about 100 mitochondria per cell and about 10 mtGenomes per mitochondrion, making mitochondrial DNA (mtDNA) more available for isolation from degraded samples relative to nuclear DNA (2–4). When mtDNA is sequenced, the focus tends to be on a region of the genome known as the mitochondrial control region (mtCR) (also known as the D-loop or hypervariable region) (5–12). In canines, the mtCR is approximately 1272 base pairs (bps) in size, is noncoding and is known to accumulate substitutions faster than any other comparably sized region of the mtGenome (13). This high rate of substitution is useful in forensic identification applications. In human investigations, the mtCR can indicate the ethnicity of a person (6). Knowing how valuable human mtDNA can be, attempts have been made to analyze mtDNA from the domestic dog for instances when dog hair is found as evidence at a crime scene (5,7,8,11,14–16). According to a 2005–2006 survey, there were then approximately 73 million domestic dogs in the United States (http://www.americanpetproducts.org/newsletter/may2005/npos.html). As dogs and humans occupy many of the same environments, it is not unexpected that dog hair is often found in criminal investigations either when a dog is directly involved in a crime or as secondary transfer from either the victim or suspect. It has been shown that while highly variable, the control region does not distinguish between dog breeds or any of the main groupings of dogs (12). In a previous study, we found that out of 552 domestic dogs, there were groups containing as many as 59 dogs of varying breeds with identical control region sequences (12). In fact, the random match probability of the mtCR for the domestic dog was found to be 4.3% as compared to between 2.5% and 0.52% for the human mtCR (4,12). Knowing that the domestic dog mtCR does not have the discriminatory power of the human mtCR, and also knowing that there are approximately 15,458 additional bps of mtGenome outside of the control region, we have sequenced the remainder of the genome for 64 domestic dogs from our mtCR study. We combined our sequences with 15 complete mtGenome sequences downloaded from Genbank (17,18). We have used phylogenetic and population genetic methods to analyze the 79 genomes and report these relationships and the variable sites in the remainder of the genome that will aid in further discriminating between dogs with common mtCR sequences.

Materials and Methods

Sample collection and DNA extraction methods were carried out as described in (12). Primers to amplify and sequence the mtGenome were designed by hand. Eleven PCR primer pairs were designed to amplify products ranging in size from 835 bp to 1918 bp. The PCR primers were designed based on the predicted sizes of the resultant amplified regions rather than based on the coordinates of a specific gene or region. This design scheme lessened our chances of amplifying mitochondrial pseudogenes, or nuclear insertions of mtDNA that are not transcribed or translated into functional proteins (19) and known to be present in canines (20). The PCR primers were also used as sequencing primers and an additional 69 sequencing primers were designed for a total of 92 primers (Table 1). Because of sequence variability, varying combinations of the 92 primers were used to sequence each dog. As a set, the complete genome primers resulted in bidirectional, overlapping, 3–4× high quality sequence coverage across the mtGenome.

[1] Animal Parasitic Diseases Laboratory, Agricultural Research Service, United States Department of Agriculture Building 1180, Beltsville, MD 20705.

[2] Molecular Methods and Subtyping Branch, Division of Microbiology, Office of Regulatory Science, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, College Park, MD 20740-3835.

*This work has been presented at The National Institute of Justice Conference 2007 and at the George Washington University Research and Discovery Day, 2007. Both instances were in poster form.

Received 15 Jan. 2008; and in revised form 11 April 2008; accepted 20 April 2008.

TABLE 1—*List of all primers (PCR and sequencing) used to sequence the canine mtGenome excluding the mtCR. The primer name, based on start coordinate relative to the Kim et al. (18) reference sequence in the 5′–3′ orientation, the primer sequence (5′–3′ orientation), the start coordinate and stop coordinate are listed.*

Primer Name	Primer 5′–3′	5′Coordinate	3′Coordinate
1620F (PCR1)	TGTTGAGCTGGAACGCTTTC	1639	1620
549F	GCTAGTAGTCCTCTGGCGAA	574	549
84F	GGTTTGCTGAAGATGGCG	701	684
1191F	GGTACTATCTCTATCGCTCC	1210	1191
16625R (PCR1)	CGCATTTGGTCTCGTAGTCT	16625	16644
171R	GGAGCAGGTATCAAGCACAC	171	190
556R	GAGGACTACTAGCAATAGCT	556	575
997R	CATACCGGAAGGTGTGCTT	997	1015
2978F (PCR2)	GTTAGGGCTAGTGATAGAGC	2997	2978
1770F	GTGGTCTATCCGTTCCTGAT	1789	1770
2400F	GGTCGTAAACCCTATTGTCG	2419	2400
1418R (PCR2)	AAGCCTAACGAGCCTGGTG	1418	1436
1999R	CGGTATCCTGACCGTGCAA	1999	2017
2512R	GGAGTAATCCAGGTCGGTTT	2512	2531
2556R	GTACGAAAGGACAAGGGATG	2556	2575
4411F (PCR3)	GTTTGATTTAGTCCGCCTCAG	4431	4411
3220F	GCGTGGATAGTGTAAATGAC	3239	3220
3804F	GGTAGCACGAAGATCTTTGA	3823	3804
3945F	GGTTCCTGTCATGATAGTTG	3964	3945
2881R (PCR3)	CCTTCAACCAATCGCAGACG	2881	2900
3479R	GCATTCCACAACCCATTCAT	3479	3498
3645R	TATGCATATGACATGTTGCC	3645	3664
4188R	CCATCGCATCCATCATGATA	4188	4207
5949F (PCR4)	GTAATTCCAGCAGCCAGTAC	5968	5949
4939F	CCTAGTCCAAGACTGATAGT	4958	4939
5407F	GGCTCATGCTCCAAATAGTA	5426	5407
5583F	GGAAACTGACTAGTGCCGTT	5602	5583
6118F	CCTGAGTAGTAAGTGACAA	6136	6118
4241R (PCR4)	CCATTCCACTTCTGAGTTCC	4241	4260
4188R	CCATCGCATCCATCATGATA	4188	4207
4274R	GGAATTACGCTCATATCAGG	4274	4293
4792R	CCTGCGACTCACATATAGCA	4792	4811
4793R	CTGCGACTCACATATAGCAC	4793	4812
5481R	GGTACTTTACTAGGTGACGA	5481	5500
7642F (PCR5)	CAATGGGTATAAAGCTGTGG	7661	7642
6352F	AAGCTCATAGCATAGCTGG	6372	6352
6415F	GGACGAATTAGCTAGGACAA	6434	6415
7035F	GAGTTGAAATGGGTACGCCA	7054	7035
5871R (PCR5)	GCAATATCCCAGTATCAAACT	5871	5891
6044R	ACACCTATTCTGATTCTTCG	6044	6063
6212R	AGCTCACCATATGTTTACCG	6212	6231
6352R	CTCCAGCTATGCTATGAGCT	6352	6371
7032R	CTATGGCGTACCCATTTCAA	7032	7054
9264F (PCR6)	GAATGTAGAGCCAATAATTACG	9285	9264
8015F	CGATCAGTACCACAATAGG	8033	8015
8152F	GAGCTCAGGTTCGTCCCTTT	8171	8152
8825F	GAATGTGCCTTCTCGGATCA	8844	8825
7512R (PCR6)	TGCATTCATGAGCCGTTCC	7512	7530
7804R	TGCCACAGCTAGATACATCC	7804	7823
8084R	CGGTTAATCTCCATTCAGCA	8084	8103
8681R	CAAGCCCATGACCGCTGACA	8681	8700
11021F (PCR7)	CTGTTTGACGGAGACAGATAG	11041	11021
9722F	TTGGTTTGTGACGCTCAGG	9740	9722
9994F	CCTCTAAGCATAGTAGCGAT	10013	9994
10625F	GTAGAGTCCTGCGTTTAGTC	10644	10625
9190R (PCR7)	GAGACATCTTTTACAATCTCCG	9190	9211
9628R	GGATCTGCTCGCCTACCTT	9628	9646
9785R	TCCTAGCTGCGAGCCTAG	9785	9802
10278R	CACGACAACATATGGTTTGC	10278	10297
10565R	TTGAAGCAACACTGATTCCG	10565	10584
12543F (PCR8)	GCGGATAAGAAGAAATACTCC	12563	12543
11508F	GCAGTAGGTGCAAGGTCATT	11527	11508
12062F	CTATGATAGACCACGTGACA	12081	12062
10844R (PCR8)	GACTACCAAAAGCACACGTAG	10844	10864
10886R	TAGTACTTGCCGCTGTACTCC	10886	10906
11270R	CCTGATGACTATTAGCAAGC	11270	11289
11892R	GCTACTTCTTACGCGTTCAT	11892	11911
11945R	CTCAGGACAGGAAACAATCA	11945	11964
13799F (PCR9)	GTTGTCTGAATTGTTGACTGC	13819	13799
12723F	GGCTGGTTAATGCCAATTGT	12742	12723

TABLE 1—*Continued.*

Primer Name	Primer 5′–3′	5′Coordinate	3′Coordinate
12730F	TAAGTAGGGCTGGTTAATGC	12749	12730
13268F	GTTCTAGTGCCAGGATGAAA	13287	13268
13565F	TAAGGATTAGTAGACTGAGG	13584	13565
12234R (PCR9)	CTACTTATTGGATGATGGTACG	12234	12255
12415R	TACTTGGCCTACTACTAGC	12415	12433
12525R	AGCACAATAGTTGTAGCAGG	12525	12544
12759R	CACATCTGCACTCACGCATT	12759	12778
13206R	ATCCCACAGATAACTATGCC	13206	13225
13352R	CCTTGGCTACTATCCAACCA	13352	13371
14810F (PCR10)	GTCTGAGTCTGATGTGATTCC	14830	14810
14030F	GCCACTAAACCATCTCCTAT	14049	14030
14253F	TCAAGCAGAGATGTTAGACG	14272	14253
14390F	CGTAGTTAACGTCTCGGCA	14408	14390
13622R (PCR10)	ATTAATAATGATCAGCCTGTAAC	13622	13644
13973R	TTCAGAACAATCGCACAACC	13973	13992
14267R	GCTTGATGGAACTTCGGATC	14267	14286
15513F (PCR11)	GAGGGGAGAAGGGTTTACC	15531	15513
14933F	TGTAGTTATCTGGGTCTCC	14951	14933
15012F	GGATCGTAGGATAGCATAGG	15031	15012
14696R (PCR11)	AAAGCAACCCTAACACGATTC	14696	14716
14933R	GGAGACCCAGATAACTACT	14933	14951
15233R	GGACAAGTCGCTTCAATCTT	15233	15252

PCR and sequencing were carried out as described in (12). Upon completion of sequencing, a check for pseudogenes was conducted. Pseudogenes are nonfunctioning and selection against mutations in the pseudogenes is not strong. As such, one way to look for potential pseudogenes is by translating the DNA into amino acid sequence and look for misplaced start or stop codons, shifted open reading frames, or difference in the amino acid composition as compared to the translation of the known mitochondrial gene sequence. The gene coding regions from each genome sequenced in this study were translated into their corresponding amino acids.

A Genbank search revealed 15 additional complete mtGenomes had been sequenced for the domestic dog. The forensic version of Sequencher 4.1.4FB19 (Gene Codes Corporation, Ann Arbor, MI) was used to edit and align all 79 mtGenome sequences. Alignments were built according to the previously defined criteria for gap placement in forensic evaluations (21). Standard IUB codes were used for polymorphic sites. A recommendation has been made to follow human mtCR methods and compare domestic dog mtCR sequences with a standard reference sequence in an effort to standardize canine mitochondrial nucleotide nomenclature (15). We continued with this recommendation by using the first published canine mtGenome as the reference mtGenome sequence (18). Using a reference sequence allows base coordinates to be compared across different studies (15), thus all coordinates mentioned in this research are in terms of the Kim et al. (18) reference sequence.

Arlequin 3.11 (22) was used to search for groups of dogs with identical mtGenome sequences, or haplotypes, and to calculate the frequency of these haplotypes. Individuals representing each unique haplotype were aligned to the reference sequence and the coordinates and base calls of the single nucleotide polymorphisms (SNPs) were recorded in an Excel spreadsheet.

Using Winclada (23), the alignment was transposed from DNA to numeric characters (A = 0, C = 1, G = 2, T = 3) using the view, numeric mode option. As with our previous control region study, Nona (24) and Winclada were used to build a phylogenetic tree to evaluate the relationships between the canines based on mtGenome sequences. A heuristic search was performed on the data following recommended search strategies (25). If the search resulted in multiple most parsimonious trees, a strict consensus tree was created. A strict consensus tree shows only those groups that exist in complete agreement among all most parsimonious trees. Upon obtaining a final tree, the relationships of the dogs were evaluated and dogs were assigned to a haplogroup based on spatial relation on the tree with other dog mtGenome sequences. Since this is the first study to identify and name haplotypes of the mtGenome, we built upon the previously established mtCR naming scheme with the intent of including the haplotype information of the entire genome, mtCR + mtGenome, into the new name. To convey the mtCR haplotype information, the mtCR haplotype name is used within the mtGenome haplotype name but modified by inserting the word "mtGenome" before the mtCR haplotype and decimal followed by a numerical distinction indicating different mtGenome types. For example, 2 individuals with the mtCR haplotype B1a but with different mtGenome haplotypes would now be called mtGenomeB1a.1 and mtGenomeB1a.2. As with the mtCR naming scheme, if an ambiguous base is present in the haplotype, the word "Ambig" is inserted into the haplotype name.

Winclada was also used to identify informative SNPs, defined as those SNPs that define a group of two or more individuals. Using the "mop informative characters/delete selected characters" function and then using the character diagnoser to trace each character on the tree, informative SNPs were identified. The length and retention index (*ri*) statistics were recorded for each informative SNP. The length is the number of times the nucleotide state at a given position changes on the tree. The *ri* is a measure of informative sites in two individuals being the result of shared common ancestry and not convergence. The *ri* scores can range from 100 to 0, a score of 100 being obtained when the character change arose only once in the evolution of the group and thus defines all members of a clade. The scores get progressively lower until a score of 0 is reached indicating all character changes arose independently.

SNPs were classified into three rankings based on the same criteria as in (12) except, due to the smaller dataset size, the third level of ranking contains informative SNPs that define groups of 8 or more individuals, or 10% of the total dogs in the dataset.

All statistics were either calculated in Arlequin or by hand. General population statistics including mean number of pairwise differences and nucleotide diversity were calculated in Arlequin on the dataset as a whole with each individual defined as a unique haplotype (not removing identical taxa) as well as by separating dogs into

TABLE 2—*List of Genbank accession number, source (publication citation) and breed sample ID of each sequence used in the current study. The breed sample ID column simply lists breed for dogs from (17,18) but dogs from (12) are listed as breed followed by a unique numerical identifier and then either a "P" or "M" representing either purebred or mixed.*

Accession Number	Source	Breed Sample ID
DQ480493	Bjornerfeldt et al., 2006	Black Russian Terrier
DQ480495	Bjornerfeldt et al., 2006	Cocker Spaniel
DQ480490	Bjornerfeldt et al., 2006	Flat Coated Retriever
DQ480489	Bjornerfeldt et al., 2006	German Shepherd
DQ480491	Bjornerfeldt et al., 2006	Irish Setter
DQ480496	Bjornerfeldt et al., 2006	Irish Soft Coated Wheaten Terrier
DQ480492	Bjornerfeldt et al., 2006	Jamthund
DQ480502	Bjornerfeldt et al., 2006	Jamthund
DQ480498	Bjornerfeldt et al., 2006	Miniature Schnauzer
DQ480494	Bjornerfeldt et al., 2006	Poodle
DQ480500	Bjornerfeldt et al., 2006	Shetland Sheepdog
DQ480499	Bjornerfeldt et al., 2006	Siberian Husky
DQ480501	Bjornerfeldt et al., 2006	Swedish Elkhound
DQ480497	Bjornerfeldt et al., 2006	West Highland White Terrier
NC_002008	Kim et al., 1998	Sapsaree
EU408245	Webb and Allard, 2009	Akita 1P
EU408246	Webb and Allard, 2009	American Cocker Spaniel 1P
EU408248	Webb and Allard, 2009	Australian Shepherd 1P
EU408249	Webb and Allard, 2009	Australian Shepherd 7P
EU408247	Webb and Allard, 2009	Australian Terrier 1P
EU408254	Webb and Allard, 2009	Basset Hound 2P
EU408255	Webb and Allard, 2009	Basset Hound 3P
EU408256	Webb and Allard, 2009	Basset Hound 4P
EU408250	Webb and Allard, 2009	Bichon Frise 3P
EU408251	Webb and Allard, 2009	Blue Heeler 1P
EU408252	Webb and Allard, 2009	Bolognese 1P
EU408253	Webb and Allard, 2009	Boxer 6P
EU408257	Webb and Allard, 2009	Brittany Spaniel 1M
EU408264	Webb and Allard, 2009	Cairn Terrier 4P
EU408260	Webb and Allard, 2009	Cardigan Corgi 2P
EU408263	Webb and Allard, 2009	Cavalier King Charles Spaniel 9P
EU408262	Webb and Allard, 2009	Chihuahua 5P
EU408261	Webb and Allard, 2009	Chihuahua 11M
EU408258	Webb and Allard, 2009	Cockapoo 1M
EU408259	Webb and Allard, 2009	Cockapoo 3M
EU408266	Webb and Allard, 2009	Cocker Spaniel 1P
EU408267	Webb and Allard, 2009	Cocker Spaniel 3P
EU408268	Webb and Allard, 2009	Cocker Spaniel 8P
EU408265	Webb and Allard, 2009	Corgi 2P
EU408270	Webb and Allard, 2009	Dachshund 4P
EU408272	Webb and Allard, 2009	Dachshund1 5P
EU408269	Webb and Allard, 2009	Doberman Pinscher 5P
EU408271	Webb and Allard, 2009	Dogue de Bordeaux 1P
EU408274	Webb and Allard, 2009	English Mastiff 3P
EU408273	Webb and Allard, 2009	English Shepherd 1M
EU408275	Webb and Allard, 2009	French Bulldog 1P
EU408277	Webb and Allard, 2009	German Shepherd 12P
EU408276	Webb and Allard, 2009	Great Dane 2P
EU408278	Webb and Allard, 2009	Great Pyrenese 1P
EU408279	Webb and Allard, 2009	Havanese 3P
EU408280	Webb and Allard, 2009	Italian Greyhound 1P
EU408281	Webb and Allard, 2009	Jack Russell 6P
EU408282	Webb and Allard, 2009	Keeshond 1P
EU408283	Webb and Allard, 2009	Keeshond 2P
EU408284	Webb and Allard, 2009	Keeshond 3P
EU408285	Webb and Allard, 2009	Labradoodle 1P
EU408286	Webb and Allard, 2009	Miniature Dachshund 2P
EU408289	Webb and Allard, 2009	Neapolitan Mastiff 1P
EU408290	Webb and Allard, 2009	Neapolitan Mastiff 2P
EU408287	Webb and Allard, 2009	Newfoundland 1P
EU408288	Webb and Allard, 2009	Norwegian Elk Hound 1P
EU408293	Webb and Allard, 2009	Pit Bull 1M
EU408291	Webb and Allard, 2009	Pomerian 2M
EU408292	Webb and Allard, 2009	Poodle 7M
EU408294	Webb and Allard, 2009	Pug 5P
EU408295	Webb and Allard, 2009	Rottweiler 1P
EU408296	Webb and Allard, 2009	Rottweiler 2P
EU408297	Webb and Allard, 2009	Schipperke 1P
EU408299	Webb and Allard, 2009	Schnauzer 4P
EU408298	Webb and Allard, 2009	Sheltie 1M
EU408300	Webb and Allard, 2009	Tibetan Mastiff 1P
EU408301	Webb and Allard, 2009	Tibetan Spaniel 1P
EU408302	Webb and Allard, 2009	Toy Poodle 3P
EU408304	Webb and Allard, 2009	Unknown 1P
EU408303	Webb and Allard, 2009	Unknown 1M
EU408305	Webb and Allard, 2009	Vizsla 2P
EU408307	Webb and Allard, 2009	Walker Hound 1P
EU408306	Webb and Allard, 2009	West Highland Terrier 4P
EU408308	Webb and Allard, 2009	Yorkie/Chihuahua 1M

purebred and mixed to look for suspected evidence of inbreeding in purebred individuals and to determine whether or not individuals labeled "purebred" and "mixed" are distinguishable at the mitochondrial sequence level. The samples were also separated by regional groupings to look for local substructure. The samples were grouped by state: California, 31; Pennsylvania, 16; Nevada, 9; Virginia, 6; Mississippi, 1; and Texas, 1. Dogs were also separated into those breeds with two or more purebred individuals to look for within breed structure: Australian Shepherd, 2; Dachshund, 2; German Shepherd, 2; Neapolitan Mastiff, 2; Poodle, 2; Jamthund, 2; Rottweiler, 2; Keeshond, 3; Cocker Spaniel, 3; Basset Hound, 3. Genetic variance was assessed using Analysis of Molecular Variance (AMOVA) with 1023 permutations to assess the significance of the variation among the various sub-divisions of the dataset. Additional statistics such as probability of exclusion, or $1 - \Sigma X_i^2$, and random match probability, or ΣX_i^2 (where X_i is the frequency of the *i*th haplotype) were calculated by hand following the arrangement of individuals with identical sequences into the same group. A gamma value, which is used to account for multiple substitutions at the same nucleotide site, was estimated by GARLI version 0.951 (26) and incorporated into Arlequin for population statistic estimations under the Tamura and Nei model of evolution (27) using AMOVA.

Results

Six hundred and ninety-eight domestic dog blood, tissue, and buccal swab samples were collected from various veterinary practices and private donors across the continental United States. Of the 698 samples collected, 426 blood and tissue samples were used for control region sequencing and analysis (12). Based on the results of the control region analysis, 64 individuals were chosen for complete genome sequencing and the sequences are available on Genbank (Table 2). These individuals were chosen based on their sharing of a mtCR haplotype with a large number of other dogs in the dataset (12) and/or if the breed-type was rare or interesting. Fifty-three of the samples came from purebred individuals and 11 were mixed breed. The 64 newly collected genomes were combined with the 15 purebred dogs downloaded from Genbank (17,18) for a final dataset of 79 domestic dogs. Table 2 lists the different breeds of dog and the number of each included in this study.

All new genomes were sequenced in their entirety and the genomes ranged in size from 15,459 bps to 15,461 bps excluding the control region. The translation of the DNA sequences into corresponding amino acids to check for pseudogenes showed that all genes translated correctly.

TABLE 3—*Informative sites for the canine mtGenome excluding the mtCR. The nucleotide coordinate relative to the Kim et al. (18) reference sequence base, the observed base, the character length (L) and character retention index (ri) are listed. Those coordinates shaded gray support groups of eight or more dogs, making them the most informative SNPs found in the current dataset.*

Base	Reference	Sample	L	ri	Base	Reference	Sample	L	ri	Base	Reference	Sample	L	ri	Base	Reference	Sample	L	ri	Base	Reference	Sample	L	ri
16	T	C	1	100	4303	A	G	2	85	8242	G	A	1	100	10776	T	C	2	83	13762	T	C	2	0
162	T	C	1	100	4360	T	C	1	100	8281	T	C	1	100	10785	A	G	2	83	13777	G	A	1	100
381	T	A	1	100	4390	T	C	1	100	8323	A	G	1	100	10863	A	G	1	100	13791	T	C	1	100
445	A	G	1	100	4466	G	A	2	66	8390	G	A	1	100	10917	G	A	1	100	14474	G	A	1	100
463	T	C	1	100	4484	G	A	1	100	8425	G	A	1	100	10992	G	A	1	100	14543	T	C	1	100
557	A	G	1	100	4503	A	G	1	100	8536	C	T	1	100	11172	A	G	1	100	14608	A	G	2	90
658	A	G	1	100	4517	G	A	1	100	8569	A	G	1	100	11176	C	T	1	100	14647	T	C	2	90
1046	G	A	1	100	4572	T	C	1	100	8670	C	T	1	100	11247	A	G	1	100	14671	G	A	1	100
1204	T	C	1	100	4591	G	A	1	100	8703	G	A	1	100	11250	T	C	1	100	14692	G	A	1	100
1351	A	G	1	100	4595	C	T	1	100	8736	T	C	1	100	11322	T	C	1	100	14800	C	T	1	100
1454	G	A	2	96	4646	T	C	1	100	8760	A	G	1	100	11400	T	C	1	100	14806	T	C	1	100
1522	G	A	2	0	4940	T	C	1	100	8764	G	T	1	100	11402	T	C	1	100	14930	T	C	1	100
1662	C	T	1	100	5009	C	T	1	100	8782	T	C	1	100	11572	A	C	1	100	14977	T	C	2	91
1689	C	T	1	100	5367	C	T	1	100	8817	A	G	1	100	11625	A	G	1	100	15185	T	C	1	100
1709	G	A	1	100	5519	C	T	1	100	8853	T	C	1	100	11657	C	A	1	100	15214	G	A	1	100
1748	T	C	1	100	5624	G	A	1	100	8877	A	G	1	100	11800	T	C	1	100	15287	G	A	1	100
1756	C	T	1	100	5855	C	T	1	100	8970	T	C	1	100	11813	A	G	1	100	15372	G	A	1	100
1766	T	C	1	100	5937	C	T	1	100	8991	A	G	1	100	11839	T	C	1	100	15435	G	A	1	100
1873	A	G	1	100	6053	C	T	1	100	9219	A	G	1	100	11897	T	C	1	100					
2185	T	C	1	100	6092	G	A	1	100	9222	C	T	1	100	11948	A	G	1	100					
2232	A	G	2	96	6257	G	A	1	100	9252	T	C	1	100	11959	C	T	1	100					
2656	G	A	1	100	6302	G	A	1	100	9708	C	T	2	83	11963	C	T	1	100					
2683	G	A	1	100	6401	C	T	1	100	9825	G	A	1	100	11984	A	G	1	100					
2812	C	T	1	100	6470	G	A	1	100	9835	A	G	1	100	12063	G	A	1	100					
2833	C	T	1	100	6518	G	A	1	100	9838	G	A	1	100	12122	C	T	1	100					
2854	A	G	1	100	6554	T	C	2	91	9865.1	-	A	3	71	12200	C	T	1	100					
2962	C	T	1	100	6629	T	C	1	100	9886	G	A	1	100	12260	A	G	1	100					
3028	A	C	1	100	6711	T	A	1	100	9896	T	C	1	100	12272	T	C	1	100					
3034	T	C	1	100	6740	G	A	1	100	10060	C	T	1	100	12330	A	G	1	100					
3196	T	C	1	100	6764	C	T	1	100	10159	C	T	1	100	12346	T	A	1	100					
3388	G	A	2	0	6767	G	A	1	100	10165	C	T	1	100	12401	T	C	1	100					
3406	C	T	2	96	6860	G	A	1	100	10195	T	C	1	100	12459	G	A	1	100					
3451	C	T	1	100	6863	C	T	1	100	10257	G	A	1	100	12636	T	C	1	100					
3465	T	C	1	100	6881	G	A	1	100	10311	C	T	1	100	12665	T	C	2	95					
3469	G	A	1	100	6967	A	G	1	100	10319	T	C	1	100	12788	T	C	1	100					
3494	T	C	1	100	7014	T	C	1	100	10346	C	T	2	75	12813	G	A	1	100					
3598	G	A	1	100	7058	T	C	1	100	10404	C	T	2	96	12818	C	T	1	100					
3628	A	G	1	100	7171	G	A	1	100	10440	T	C	1	100	12968	G	A	1	100					
3937	C	T	1	100	7186	C	A	1	100	10533	A	T	1	100	13102	T	C	1	100					
3940	C	T	1	100	7450	C	T	1	100	10542	A	G	1	100	13112	G	A	1	100					
3950	A	G	1	100	7593	T	C	1	100	10557	C	T	1	100	13261	C	T	1	100					
4135	C	T	1	100	7923	T	C	1	100	10611	A	T	2	91	13426	C	T	1	100					
4169	A	G	1	100	8101	G	A	1	100	10613	A	G	1	100	13594	G	A	2	95					
4204	G	A	1	100	8108	C	T	1	100	10680	C	T	1	100	13618	A	G	1	100					
4234	C	T	1	100	8221	A	C	1	100	10725	T	C	1	100	13660	C	T	1	100					
4277	A	G	1	100	8225	T	C	1	100	10773	T	C	1	100	13708	C	T	1	100					

Following the separate alignments of each unique genome sequence to the Kim et al. (18) reference sequence, six gaps were inserted into the matrix: 1493.1, 2679.1, 7015.1, 9865.1, 9914.1 and 9914.2 and the final multiple alignment matrix size was 15,463 bps by 79 dogs.

Within the roughly 15,460 bases of the mtGenome excluding the mtCR, 356 SNPs were found (2.3%). Of the 356 SNPs, 57% ($n = 202$) were found to be informative and 26% ($n = 94$) were found to be highly informative by defining groups of eight or more dogs (approximately 1% of the dataset) (Table 3). In other words, over 1/3 of the SNPs (43%) are variations unique to an individual. Comparatively, 9.5% of 987 mtCR bases were found to be variable with 42% being unique to an individual.

TABLE 4—*This table lists the haplotype name in the leftmost column, followed by the number of dogs that possess the haplotype followed by the variable positions that define the haplotype.*

Table 4

* = autopomorphy

			*		*					*		*	*	*	*		*	*				*		*					*						*		*		*							*	*		*	*							
																1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	# of			1	1	3	4	4	5	6	6	7	7	7	9	0	0	1	2	3	4	4	5	5	6	6	7	7	7	7	7	8	1	2	6	6	6	6	8	8	8	8	9	0	0	0	1	1	2	3	3	4	4	4	4	4	5
	Dogs	1	7	6	6	8	4	6	5	0	5	3	6	6	4	4	7	9	0	5	5	9	2	9	6	8	0	4	5	5	6	7	8	3	4	5	7	8	0	1	3	5	6	2	3	4	0	9	7	0	8	0	5	6	6	9	9
	per	6	0	2	5	1	5	3	7	7	8	4	7	8	7	6	6	0	4	1	4	3	2	1	2	9	9	8	1	6	6	3	5	2	6	6	9	3	0	2	3	4	2	8	4	6	6	6	8	1	8	6	1	5	9	4	8
	haplo-																					1															1																				
Haplotype	group	T	T	T	T	T	A	T	A	T	A	T	G	A	C	G	T	T	T	A	G	-	G	T	C	C	G	T	T	C	T	A	T	A	T	G	-	G	C	C	C	A	C	A	T	G	A	T	G	A	G	C	C	T	G	T	G
mtGenomeA19a.1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA18d.1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA27c.1	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA11e.1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA26a.1	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	-	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA2b.1	2	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	T	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA17a.1	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA19a.2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA20c.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA18d.2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA18d.4	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA18d.3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA18b.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA11e.2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA11e.3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
mtGenomeA26a.2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	-	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
mtGenomeA108.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		A	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	T	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA22a.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	C	.	.	.	.	.	G	A	.	.	T	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA22a.2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	C	.	.	.	.	.	G	A	.	.	T	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA2a.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA2b.2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	T	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA16a.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA16a.2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA71.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA71.2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA16a.3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA17a.2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA17a.3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA17a.4	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
mtGenomeA98.1	1	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.		A	.	.	.	.	.	.	.	.	.	.	G	.	.	G	A	.	.	.	.	T	.	.	A	.	C	.	.	.	T	.	.	.	.	.
mtGenomeA29b.1	1	.	.	.	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	T	.	.	.	T	.	.	.	G	C	.	.	.	T	.	.	.	.	.
AmbigmtGenomeA18d.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AmbigmtGenomeA20b.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AmbigmtGenomeA18d.2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AmbigmtGenomeA11Ambig2.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AmbigmtGenomeA17a.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.
AmbigmtGenomeA97.1	1	.	.	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.	G	.	.	G	A	.	.	.	.	T	.	.	.	.	C	A	.	.	T	.	.	.	.	.
mtGenomeB1a.1/B1g.1/B1h.1	3	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeB1a.2/B1Ambig4.1	2	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeBAmbig11.1	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeB1a.3	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeB1a.4	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeBAmbig4.1	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeB1a.5	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeB10a.1	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	A	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	A	T	.	.	A	.	A
mtGenomeB6a.1	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeB6a.2	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeB30.1	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeB28.1	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeBAmbig12.1	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeB1Ambig1.1	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeB1Ambig4.1	1	C	.	.	.	A	G	.	.	.	.	C	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeB1a.6	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
AmbigmtGenomeB1a.1	1	C	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A		.	.	.	.	.	.	.	T	.	.	C	G	.	.	G	A	.	T	.	.	T	C	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeC8a.1	1	C	.	.	.	A	.	.	G	.	.	.	.	.	.	A	.	.	C	.	A		.	.	.	.	A	C	.	T	.	.	.	G	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeC12.1	1	C	.	.	.	A	.	.	G	.	.	.	.	.	.	A	.	.	C	.	A		.	.	.	.	A	C	.	T	.	.	.	G	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	T	.	.	A	.	A
mtGenomeC3a.1	1	C	.	.	.	A	.	.	G	C	.	.	.	.	.	.	.	.	C	.	A		.	.	.	.	A	C	.	T	.	.	.	G	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	T	.	.	A	C	A
mtGenomeC3b.1	1	C	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	C	.	A		.	.	.	.	A	C	.	T	.	.	.	G	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	A	C	A
mtGenomeC3b.2	1	C	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	C	.	A		.	.	.	.	A	C	.	T	.	.	.	G	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	T	.	.	A	C	A
mtGenomeC3a.2	1	C	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	C	.	A		.	.	.	.	A	C	.	T	.	.	.	G	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	A	T	.	.	A	C	A
mtGenomeC3Ambig1.1	1	C	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	C	.	A		.	.	.	.	A	C	.	T	.	.	.	G	.	.	G	A	.	T	.	.	.	.	.	.	.	C	.	.	.	T	.	.	A	C	A
AmbigmtGenomeCAmbig1.1	1	C	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	C	.	A		.	.	.	.	A	C	.	T	.	.	.	G	.	.	G	A	.	.	.	.	T	.	.	.	.	C	.	.	.	T	.	.	A	C	R
mtGenomeD1a.1	1	.	C	C	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	A		.	.	.	T	.	.	.	.	.	G	.	G	.	.	G	A	.	.	.	G	T	.	C	.	.	C	.	.	.	T	T	C	A	.	A
mtGenomeD1b.1/D2.1	2	.	.	C	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	A		.	.	.	T	.	.	.	.	.	G	.	G	.	A	G	A	.	.	.	G	T	.	C	.	.	C	.	.	.	T	T	C	A	.	A

Some haplotypes have 2 names because haplotype name represents the combined mtCR + mtGenome haplotypes for the individual. The row at the top shows the coordinate of each SNP relative to the Kim reference sequence, whose nucleotides are listed immediately below the coordinates at the varying sites. All SNPs are listed as the variable nucleotide at the corresponding position. An asterisk (*) above a coordinate indicates an informative SNP in Table 3. A dot (.) indicates a match to the reference sequence and a blank cell indicates no variation between the sample and the reference sequence.

A complete list of haplotypes can be found in Table 4 and the frequency of each haplotype as well as each dog possessing a given haplotype can be found in Table 5.

Haplogroup A was the largest group containing 60.75% (*n* = 48) of the total individuals in the dataset. Within group A there were seven groups of individuals sharing a haplotype, 25 haplotypes unique to an individual and six individuals with ambiguous base calls that could not be placed within a haplotype group. Haplogroup B was the second largest group of dogs containing 25.3% (*n* = 20) of all individuals. Of the 20 individuals only two groups were

*			*	*				*				*				*											*			*	*	*	*	*		*		*		*	*		*		*	*	*		*		*				*				*			*		*						
3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
5	6	6	7	8	9	9	9	9	1	1	2	2	2	2	3	3	3	3	4	4	5	5	5	5	5	6	8	9	0	0	0	0	1	2	3	4	5	5	6	6	6	8	9	9	9	9	0	0	0	0	2	2	3	4	4	4	5	5	5	6	7	7	7	7	7	7	8	8	8	9
9	2	7	4	0	3	4	5	6	3	6	0	2	3	7	0	4	6	9	6	8	0	1	7	9	9	4	8	4	0	3	3	4	2	5	6	4	1	6	2	6	9	5	2	3	7	7	4	5	6	9	3	5	0	0	0	7	1	5	7	2	1	3	4	4	6	6	6	6	8	6
9	8	6	4	4	7	0	0	4	5	9	4	9	4	7	3	8	0	0	6	4	3	7	2	1	5	6	0	0	9	0	5	2	6	9	7	4	9	4	4	3	9	5	5	7	1	2	1	3	5	2	6	7	2	1	7	0	8	4	2	9	1	1	0	3	4	7	0	3	1	7
T	A	C	A	T	C	C	A	C	C	A	G	C	C	A	A	C	T	T	G	G	A	G	T	G	C	T	G	T	C	T	T	G	C	A	C	T	C	C	G	T	G	C	G	C	T	A	T	C	A	G	A	G	A	C	A	G	G	T	T	T	T	G	G	G	C	G	G	C	G	A
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	T	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	T	.	.	.	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	A	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	C	.	T	.	.	.	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	T	.	.	.	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	C	.	G	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	T	C	.	.	.	.	A	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	C	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	R	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	C	.	T	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	A	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	C	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	A	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	A	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	C	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	A	.	A	A	.	.	.	T	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	A	.	A	.	.	.	.	T	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	T	.	.	.	.	G	A	.	.	.	.	.	.	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	A	.	A	.	.	.	.	T	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	A	.	A	.	.	.	.	T	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	A	.	A	.	.	.	.	T	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.	A	.	A	.	.	.	.	T	.	.
.	G	.	.	.	T	T	G	.	.	.	.	.	.	.	.	.	.	.	A	A	.	A	.	A	T	.	.	.	T	.	.	.	.	.	T	C	T	.	.	.	.	.	.	T	.	.	.	T	G	A	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	A	A	.	A	.
.	G	.	.	.	T	T	G	.	.	.	.	.	.	.	.	.	.	.	A	A	.	A	.	A	T	.	.	.	T	.	.	.	.	.	T	C	T	.	.	.	.	.	.	T	.	.	.	T	G	A	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	A	A	.	A	.

formed, 14 individuals had unique mtGenome sequences and one individual was ambiguous. Haplogroup C was the third largest group with 10.1% ($n = 8$) of all individuals. Seven of the eight individuals had unique haplotypes and one individual was ambiguous. Haplogroup D was the smallest group containing only 3.8% ($n = 3$) of all individuals and contained one group of two dogs sharing a haplotype and one individual with a unique haplotype. Figure 1 shows the distribution of individuals relative to their haplotype.

Twenty-four of the 79 dogs were identified as being identical to at least one other dog in the dataset based on mtGenome excluding

TABLE 5—*The haplotype distribution of all individuals in the dataset. Haplotype, mtCR haplotype, breed, the number of individuals per breed ([n] per breed) sharing the haplotype, the total number of individuals sharing the haplotype (Total [n]), and frequency of haplotype (%) are listed. MtCR haplotypes can be found in (12). Samples with mtCR haplotypes marked with an asterisk (*) are from Bjornerfeldt et al. (17) and are not presented in (12). The haplotype names were formed via a concatenation of the mtCR and mtGenome haplotypes. The mtGenome haplotypes are listed in Table 4.*

Haplotype	mtCR Haplotype	Breed Sampled ID	(*n*) per breed	Total (*n*)	%
mtGenomeA2a.1	A2a	West Highland White Terrier (DQ480497)	1	1	1.27
mtGenomeA2b.1	A2b	Great Dane 2P	1	2	2.53
	A2b	Schnauzer 4P	1		
mtGenomeA2b.2	A2b	French Bulldog 1P	1	1	1.27
mtGenomeA11e.1	A11e	Rottweiler 1P	2	2	2.53
	A11e	Rottweiler 2P			
mtGenomeA11e.2	A11e	Miniature Dachshund 3P	1	1	1.27
mtGenomeA11e.3	A11e	Australian Shepherd 7P	1	1	1.27
AmbigmtGenomeA11Ambig2.1	A11Ambig2	Cocker Spaniel 1P	1	1	1.27
mtGenomeA16a.1	A16a	Brittany Spaniel 1M	1	1	1.27
mtGenomeA16a.2	A16a	Italian Greyhound 1P	1	1	1.27
mtGenomeA16a.3	A16a	English Mastiff 3P	1	1	1.27
mtGenomeA17a.1	A17a	Boxer 6P	1	3	3.80
	A17a	Dogue de Bordeaux 1P	1		
	A17a	Miniature Schnauzer (D480498)	1		
mtGenomeA17a.2	A17a	Unknown 1P	1	1	1.27
mtGenomeA17a.3	A17a	Cavalier King Charles Spaniel 9P	1	1	1.27
mtGenomeA17a.4	A17a	Bichon Frise 3P	1	1	1.27
AmbigmtGenomeA17a.1	A17a	Pug 5P	1	1	1.27
mtGenomeA18b.1	A18b	American Cocker Spaniel 1P	1	1	1.27
mtGenomeA18d.1	A18d	Jack Russell 6P	1	2	2.53
	A18d	Sheltie 1M	1		
mtGenomeA18d.2	A18d	Dachshund 15P	1	1	1.27
mtGenomeA18d.3	A18d	Vizsla 2P	1	1	1.27
mtGenomeA18d.4	A18d	Cocker Spaniel (DQ480495)	1	1	1.27
AmbigmtGenomeA18d.1	A18d	Cockapoo 3M	1	1	1.27
AmbigmtGenomeA18d.2	A18d	Toy Poodle 3P	1	1	1.27
mtGenomeA19a.1	A19a	Dachshund 4P	1	2	2.53
	A19a	German Shepherd 12P	1		
mtGenomeA19a.2	A19a	Sapsaree (NC_002008)	1	1	1.27
mtGenomeA19a.2	A19a	Australian Shepherd 1P	1	1	1.27
AmbigmtGenomeA20b.1	A20b	English Shepherd 1M	1	1	1.27
mtGenomeA20c.1	A20c	Chihuahua1 1M	1	1	1.27
mtGenomeA22a.1	A22a	Neopolitan Mastiff 1P	1	1	1.27
mtGenomeA22a.2	A22a	Neopolitan Mastiff 2P	1	1	1.27
mtGenomeA26a.1	A26a	West Highland Terrier 4P	1	3	3.80
	A26a	Cairn Terrier 4P	1		
	A26a	Irish Soft Coated Wheaton Terrier (DQ480496)	1		
mtGenomeA26a.2	A26a	New Foundland 1P	1	1	1.27
mtGenomeA27c.1	A27c	Keeshond 1P	3	3	3.80
	A27c	Keeshond 2P			
	A27c	Keeshond 3P			
mtGenomeA29b.1	A29b*	Siberian Husky (DQ480499)	1	1	1.27
mtGenomeA71.1	A71	Corgi 2P	1	1	1.27
mtGenomeA71.2	A71	Akita 1P	1	1	1.27
AmbigmtGenomeA97.1	A97	Tibetan Mastiff 1P	1	1	1.27
mtGenomeA98.1	A98	Chihuahua 5P	1	1	1.27
mtGenomeA108.1	A108*	Irish Setter (DQ480491)	1	1	1.27
mtGenomeBAmbig4.1	BAmbig4	Doberman Pinscher 5P	1	1	1.27
mtGenomeBAmbig11.1	BAmbig11	Unknown 1M	1	1	1.27
mtGenomeBAmbig12.1	BAmbig12	Yorkie/Chihuahua 1M	1	1	1.27
mtGenomeB1Ambig1.1	B1Ambig1	Australian Terrier 1P	1	1	1.27
mtGenomeB1Ambig4.2	B1Ambig4	Cardigan Corgi 2P	1	1	1.27
mtGenomeB1a.1	B1a	Labradoodle 1P	1	3	3.80
mtGenomeB1g.1	B1g*	Shetland Sheepdog (DQ480500)	1		
mtGenomeB1h.1	B1h*	Poodle (DQ480494)	1		
mtGenomeB1a.2	B1a	Basset Hound 4P	2	2	2.53
mtGenomeB1Ambig4.1	B1Ambig4	Basset Hound 2P			
mtGenomeB1a.3	B1a	Tibetan Spaniel 1P	1	1	1.27
mtGenomeB1a.4	B1a	Bolognese 1P	1	1	1.27
mtGenomeB1a.5	B1a	Poodle 7M	1	1	1.27
mtGenomeB1a.6	B1a	Great Pyrenese 1P	1	1	1.27
AmbigmtGenomeB1a.1	B1a	Basset Hound 3P	1	1	1.27
mtGenomeB6a.1	B6a	Walker Hound 1P	1	1	1.27
mtGenomeB6a.2	B6a	Schipperke 1P	1	1	1.27
mtGenomeB10a.1	B10a	Cocker Spaniel 8P	1	1	1.27
mtGenomeB28.1	B28	Cockapoo 1M	1	1	1.27
mtGenomeB30.1	B30*	Flat Coated Retreiver (DQ480490)	1	1	1.27

TABLE 5—*Continued.*

Haplotype	mtCR Haplotype	Breed Sampled ID	(*n*) per breed	Total (*n*)	%
AmbigmtGenomeCAmbig1.1	CAmbig1	Blue Heeler 1P	1	1	1.27
mtGenomeC3Ambig1.1	C3Ambig1	Cocker Spaniel 3P	1	1	1.27
mtGenomeC3a.1	C3a	Pomerian 2M	1	1	1.27
mtGenomeC3a.2	C3a	Havanese 3P	1	1	1.27
mtGenomeC3b.1	C3b*	Black Russian Terrier (DQ480493)	1	1	1.27
mtGenomeC3b.2	C3b*	Swedish Elkhound (DQ480501)	1	1	1.27
mtGenomeC8a.1	C8a	Pit Bull 1M	1	1	1.27
mtGenomeC12.1	C12*	German Shepherd (DQ480489)	1	1	1.27
mtGenomeD1a.1	D1a	Norweigian Elk Hound 1P	1	1	1.27
mtGenomeD1b.1	D1b*	Jamthund (DQ480502)	2	2	2.53
mtGenomeD2.1	D2*	Jamthund (DQ480492)			

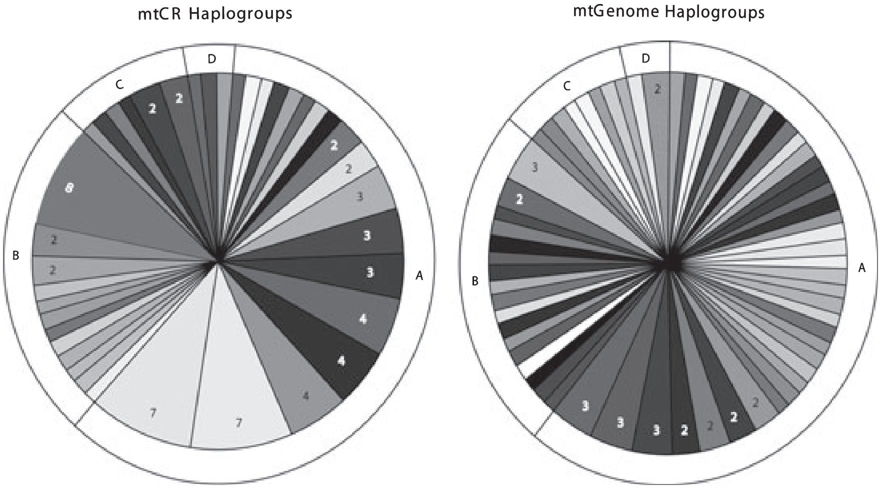

FIG. 1—*Distribution of haplotypes. Pie charts showing distributions of individuals, excluding those with ambiguous base calls that share identical DNA sequence, or haplotypes. The chart on the left presents mtCR haplotypes and the chart on the right mtGenome haplotypes for the same set of dogs. Regardless of mtCR or mtGenome sequence, the trend of haplogroup A containing the most dogs followed by haplogroups B, C, and then D is retained. The numbers inside of the slices represent the number of individuals found with that particular haplotype. Haplogroup B has the largest single instance of individuals with the same haplotype (*n *= 8) for the mtCR dataset. For the mtGenome dataset, the largest groups contain three individuals and are found in both A and B.*

mtCR sequence. There was one instance of a purebred and a mixed breed dog sharing an identical sequence and the remaining instances of shared sequences all occurred within purebred dogs. None of the dogs evaluated were identical to the Kim et al. (18) reference sequence. Of the unique haplotypes, eight of those were due to individuals having ambiguous base calls in their sequence. Excluding these eight sequences from the calculations, 66.2% of the mtGenomes excluding mtCR sequenced were unique in the dataset of 71 dogs. This is much higher than the 18.3% unique canine mtCR haplotypes found in our previous study of 552 mtCRs. When considering only the mtCRs of the 79 dogs used in the current study, excluding those dogs with ambiguous mtCR base calls ($n = 9$), 52 dogs were identical to at least one other dog in the dataset, or only 25.7% ($n = 18$) of the mtCR sequences were unique (Figs. 1 and 2, Table 5).

When assessing the same set of dogs for the two different mitochondrial regions the phylogenetic relationships were highly similar. When using mtGenome sequence excluding the mtCR all individuals formed groups with the same individuals as they did using mtCR sequence alone (Fig. 3).

A mutational "hotspot" has been reported in the canine mtCR (28) and confirmed (12). In the most recent study, this hotspot was defined by 22 mutations occurring in a region of 60 bps, or 1 mutation in every 2.7 bases, as opposed the calculated average rate of 1 mutation in every 15 bases for the mtCR. In the mtGenome, the calculated average mutation frequency is 1 mutation in every 50 bases. Looking at the distribution of mutation within the mtGenome, there are clusters of sequence variation and stretches of the genome where no SNPs are found. The regions with some of the highest frequency of SNPs were bases 10,251–10,354 with 9 SNPs in 103 bases, 11,800–12,006 with 16 SNPs in 206 bases, and bases 8661–9028 with 23 SNPs in 367 bases. The frequency of SNPs in these three regions is 1 in 11.5, 1 in 13 and 1 in 16, respectively. While this is not close to the 1 in 2.7 frequency of the mtCR hotspot, it is significantly greater than the 1 in 50 mutation rate mtGenome average. Conversely, there were regions of 400 base pairs or larger that had very few SNPs. The regions spanning 1767–2645 (878 bp) and 9220–9824 (604 bp) have only three SNPs and the region spanning 13,792–14,328 (536 bp) has only two SNPs. The largest region without any SNPs occurs between bases 9253–9707. This 454 bp region, as well as the larger 604 bp region with only three SNPs in which it's contained, spans the coding region for the end of COIII gene, the tRNA-Gly and the beginning of the ND3 gene. Likewise, the other regions with only a few SNPs span the coding region for 16S rRNA and the coding region of the ND6 gene, tRNA-Glu, and the CYTB gene.

Based upon the frequency of each haplotype, the random match probability for the mtGenome dataset as a whole was calculated to

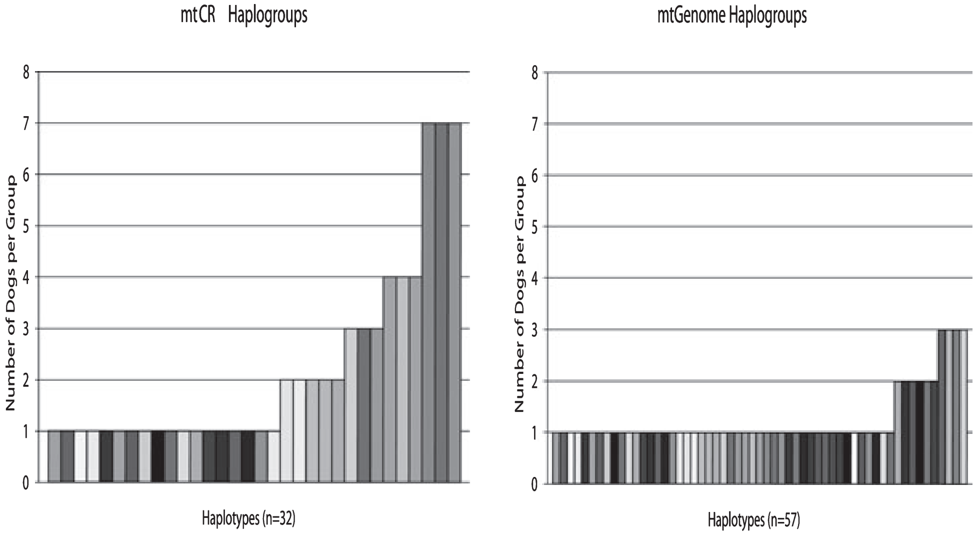

FIG. 2—*Distribution of haplotypes based on group size. These two graphs show a comparison of mtCR and mtGenome haplotype groups. The haplotypes are represented along the x-axis and the number of dogs sharing a particular haplotype represented by the y-axis. The graphs show that the mtGenome has more individuals with unique haplotypes and fewer groups of two or more identical samples compared with the mtCR for the same 79 dogs. Dogs from each dataset with ambiguous base calls were not included (mtCR,* n = 9*), (mtGenome,* n = 8*).*

be 0.018 and the probability of exclusion was calculated to be 0.982. This implies that 98 individuals of 100 can be excluded based on the mtGenome dataset, or that 2 of 100 individuals may have identical haplotypes simply by chance. Comparatively, the random match probability for the mtCR was calculated to be 0.041 with 96 of 100 individuals excluded based on the mtCR dataset.

Using GARLI, an alpha value for the gamma correction to account for multiple substitutions at a single nucleotide site was calculated to be 0.0087, which was rounded to 0.01. Treating all newly collected sequences as a single population, the mean number of pairwise differences was 84.14 ± 36.58 and the nucleotide diversity was 0.005441 ± 0.002621. When the population was split into purebred and mixed breed individuals the mean number of pairwise differences decreased slightly though not significantly to 83.20 ± 36.24 for purebred and increased for mixed breed to 90.12 ± 42.05. The nucleotide diversity also decreased slightly to 0.005380 ± 0.002598 for purebred and increased for mixed breed to 0.005829 ± 0.003069.

The fixation index (Φst) values in Table 6, which represent the proportion of genetic variation within a subpopulation relative to the total population, are very low for the purebred versus mixed breed values and geographic state of origin comparison, showing that grouping dogs by these factors has no genetic basis. As can be seen in Table 5, dogs of the same breed do group together in some instances, but there are also cases, such as the cocker spaniels, where dogs of the same breed are spread out across the three different haplogroups. The AMOVA shows that almost 30% of the variation can be attributed to among breed variation and the *p*-value, estimated by 1023 permutations, demonstrates the significance of the results ($p < 0.05$).

Discussion

The aim of this study was to sequence the mtGenome from multiple domestic dogs to search for informative SNPs that would more fully resolve the large haplotype groups formed by using the mtCR sequence alone, and to assess the utility of the mtGenome for forensic analyses. Individuals were chosen for mtGenome sequencing because either they belonged to one of the large mtCR haplotype groups or the breed was of interest. The 64 newly sequenced mtGenomes combined with the 15 mtGenomes downloaded from Genbank form the largest domestic dog mtGenome dataset to be published to date and the first to be used to identify domestic dog mtGenome haplotypes.

During sample collection, donors were asked to determine breed and breed type (either purebred or mixed). As the authors never saw the actual dog, breed and type were never changed, even when the declarations were questionable. For example, two samples were received with one being labeled "West Highland White Terrier" and the other "West Highland Terrier." While these two dogs could very well be of the same breed, they were distinguished as different breeds in the current dataset based on the differing donor descriptions. Individuals with unknown breed or breed type were considered mixed unless otherwise listed by the donor.

When comparing the mtGenome excluding the mtCR to the mtCR, it was revealed that while the mtGenome has more haplotypes, the mtCR has a higher overall percentage of SNPs. Also, the percentage of SNPs unique to an individual is about the same for the two datasets. While it may seem counter-intuitive that such a comparatively small region would have a higher percentage of SNPs, it must be remembered that the mtCR is noncoding, meaning it is not translated into an amino acid sequence and therefore lacks this kind of biological constraint to prevent nucleotides from mutating. The majority of the mtGenome excluding the mtCR codes for RNAs or proteins with important biological functions, making the probability of a SNP occurring in one of those regions much lower (13). When SNPs do occur in a coding region, it is more likely that they are unique or possessed by only a small number of individuals, leading to more haplotypes with unique SNPs or unique combinations of SNPs within the mtGenome, which is what we see in our dataset. Collectively, our results show that while there is more variability in the mtCR, the percentage of unique SNPs is relatively constant throughout the genome. Incorporation of SNPs outside of the mtCR increases the number of informative SNPs for forensic use to 57% of the total SNPs found.

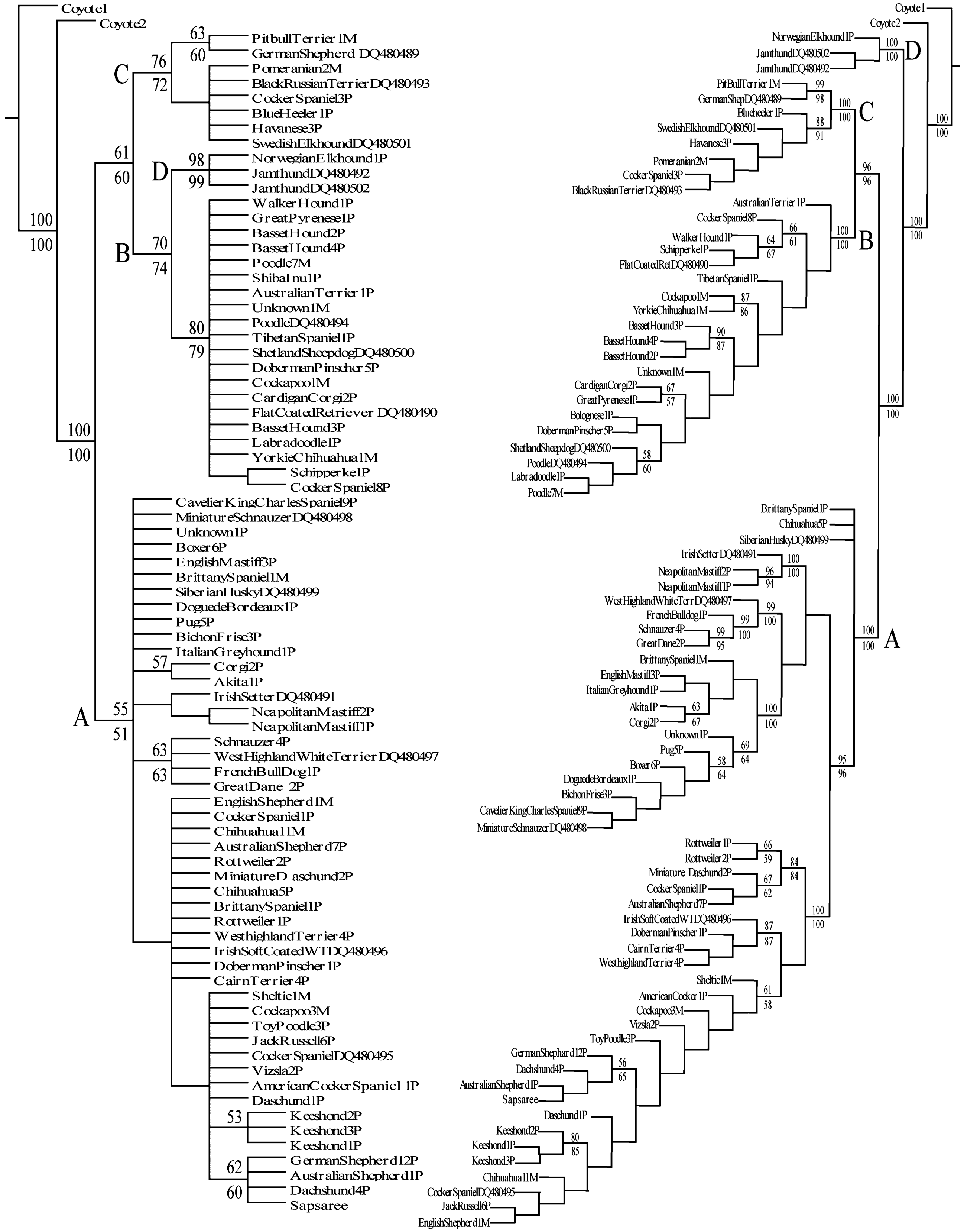

FIG. 3—*mtCR and mtGenome Phylogenetic trees. Parsimony reconstructions of the 79 dogs and 2 coyotes using only mtCR sequences (left) or only mtGenome sequences (right). Each tree is a strict consensus of all equally parsimonious solutions. Two hundred and sixteen equally likely mtCR trees were found with lengths of 106, CI of 70, and RI of 94. Two equally likely mtGenome trees were found with lengths of 995, CI of 94, and RI of 97. The letters "A," "B," "C," and "D" represent the previously identified major haplogroup labels. Bootstrap support scores >50 are shown above the branches, jackknife support scores >50 below. While the relationships of the major haplogroups changes, and the order of the dogs within the groups changes, close inspection of each major group will show that the same dogs fall within the same groups regardless of the region of DNA sequence being used.*

TABLE 6—*Results of three separate AMOVAs.*

Dataset	Source of Variation	Degrees of Freedom	Percentage of Variation
Purebred vs mixed	Among breed groups	1	0.20
	Within breed groups	77	99.80
	Total	78	100
		Φst = 0.00198, *p* = 0.33	
By state*	Among states	5	0 (−4.72)
	Within states	58	104.72
	Total	63*	100
		Φst = 0 (−0.04720), *p* = 0.86	
By breed	Among breeds	9	66.06
	Within breeds	13	33.94
	Total	22	100
		Φst = 0.66064, *p* = 0.00	

Purebred versus mixed specific Φst indices	
mtGenome purebred only	0.00198
mtGenome mixed only	0.00201
State specific Φst indices	
Pennsylvania	−0.05990
California	−0.04644
Nevada	−0.05818
Virginia	−0.04122
Mississippi	0.07422
Texas	0.07422

Breed specific Φst indices

Pop#	Name	Φst
1	Australian Shepherds	0.75605
2	Basset Hounds	0.80214
3	Cocker Spaniels	0.10163
4	Dachshunds	0.78622
5	German Shepherds	0.27341
6	Neapolitan Mastiffs	0.79376
7	Poodles	0.80130
8	Jamthunds	0.80884
9	Rottweilers	0.80884
10	Keeshonds	0.80884

Using the entire dataset, dogs were sorted as purebred or mixed. The percent variation among versus between these breed types as well as the degrees of freedom for each grouping are listed. Using all dogs except the Kim et al. (18) reference sequence or the 14 samples from Bjornerfeldt et al. (17), the genetic variation was assessed among all dogs treated as one population versus each state being treated as an individual population (Within States). Using only those dogs that belonged to breeds with >6 members present in our dataset dogs were sorted by breed. Due to the decreased dataset size, the degrees of freedom values for the By States and By Breed analyses are less than the purebred versus mixed analysis. For each datasets, Φst was estimated for the among population variation as well as each of the dog groupings. The significance, reported as a *p*-value, was derived from 1023 permutations.

Collectively, the 79 dogs in our dataset formed 10 groups and 47 unique haplotypes with 8 ambiguous sequences. The ambiguous base calls were due to true polymorphisms within the individual dog samples due to the multiple genomes per cell (2,3). While the number of individuals with unique haplotypes may seem high, it is important to keep in mind that this is the first study of its kind, and the number will almost surely decrease as more dog mtGenomes are evaluated. Relative to the mtCR, this number will likely always be higher due to larger region and higher constraints against mutation on the coding portions of the mtGenome.

As mentioned above, the number of individuals that share identical mtGenome sequences is smaller than the number of individuals that share mtCRs for the same dogs (Figs. 1 and 2). This illustrates how the additional sequence variation of the mtGenome can be used to add phylogenetic resolution to large groups that often result from mtCR sequencing. Figure 1 shows how the dogs are situated relative to their haplotype. Of note is the single instance within haplogroup D where the mtCR sequences provide unique sequence variation for dogs that possess identical mtGenomes. This shows that ideally, one should sequence the entire mtGenome, including the mtCR, to fully utilize the DNA sequence variability within this genome. Figure 2 demonstrates the phenomenon that was seen in our larger mtCR study: while there are many canine mitochondrial control region haplotypes, most dogs share the common types while the minority of dogs have unique or rare types. The distribution of the dogs within the mtGenome haplotype groups shows that the additional variation found in the remainder of the mtGenome provides, in most cases, resolution of the large groups formed by mtCR sequences alone.

The distributions of dogs within each haplogroup were consistent with the mtCR groupings. As previously reported, when using only the mtCR sequence group A contained the most individuals followed by groups B, C, and D (12). When evaluating the mtGenome groups in the same manner, the same trend persists. Group A had the most individuals followed by B, C, and D. When viewing the relationships of the dogs in the trees shown in Fig. 3, it can be seen that not only do the sizes of the groups correspond between datasets, but also the members of each group. Dogs that grouped together based upon their mtCR also grouped together based upon their mtGenome excluding the mtCR sequences, indicating that the signal present in the mtCR is also present in the remainder of the mtGenome. This result is expected as the mtGenome does not undergo recombination and as such acts as a single locus. This is promising for forensic use of canine mitochondrial DNA as it shows that the entire mtGenome can be used to identify samples because the results from different regions of the genome do not conflict.

The importance of the mutational "hot spots" within the mtGenome is that forensic samples are often degraded, making it difficult to obtain complete sequence through large areas. Also, the mtGenome is 92% larger than the mtCR and as such it is much more expensive to sequence. By identifying the most variable regions, we have provided coordinates where future groups can focus DNA sequencing efforts. Conversely, the regions where no SNPs were found could be avoided.

The random match probability results show that when considering the remainder of the mtGenome, there is a lower chance of a random match compared to using the mtCR alone. This is significant as it shows that the probability of finding a coincidental match when using the mtGenome is lower than when using the mtCR alone.

The results of the pairwise difference and nucleotide diversity assessments are consistent with the findings of the mtCR study. Although not statistically significant, they indicate that mixed breed dogs come from a more variable gene pool and, as expected, have more genetic variation than purebred dogs. The ancestral lines of purebreds should contain only the DNA of individuals from the same breed or the founding breeds, resulting in more constrained physical as well as genetic characteristics.

As we never actually saw the dogs from which our samples were obtained, we wanted to test the significance of the purebred versus mixed labels. Our results agree with the nucleotide diversity results showing that there is not a significant amount of genetic variation between the group of dogs labeled "purebred" and those dogs labeled "mixed." This illustrates that not knowing whether a dog is purebred or mixed has very little consequence on the dataset in terms of mtDNA. Additionally, we show that geographic location of sample collection is not relevant when evaluating dogs from the continental United States via mtGenome haplotypes. Conversely, the AMOVA results are significant when dogs are grouped based on breed demonstrating that dogs of the same breed, while perhaps not possessing identical mtGenome sequences, have more similar sequence composition than expected by chance. The AMOVA results support our previous mtCR dataset findings allowing us to draw the same conclusions. First, classifying breeds by breed type (purebred or mixed) is trivial when it comes to mtDNA. Second, there is no need for local canine mitochondrial SNP databases. Finally, there is some degree of population substructure when dogs are grouped by breed. This is most likely due to the higher amounts of inbreeding of purebred dogs, underscoring the need to collect multiple individuals from the same breed in the construction of a mitochondrial SNP database.

In summary, consistent with the mtCR results, analysis of the SNPs in the remainder of the mtGenome does not group dogs by breed or any other common domestic dog grouping. However, the SNPs found in the remainder of the mtGenome are useful in that they provide additional discriminatory sites that resolve common mtCR haplotype groups. Within our dataset of 79 domestic dog mtGenomes excluding the mtCR, 2.3% of the nucleotides were found to be variable. Fifty-seven percent of the variable sites were informative by supporting groups of two or more dogs and 26% of the informative sites were highly informative by supporting groups of eight (1%) or more dogs. When comparing haplotype groups formed from the mtCR sequences alone and the mtGenome sequences without the mtCR for the same set of 79 dogs, it becomes obvious that the SNPs found in the remainder of the mtGenome have a higher discriminatory power overall. When looking at the mtCR alone and excluding ambiguous sequences, there are 18 individuals (25.7%) with unique mtCR sequences and 52 dogs (74.3%) forming 14 groups with up to seven dogs per group. Comparatively, when looking at the same 79 dogs using mtGenome sequences without the mtCR and excluding ambiguous sequences, the distribution shifts with 24 dogs (33.8%) forming 10 groups containing at most three dogs and the remaining 66.2% ($n = 47$) of the dogs having unique haplotypes. While, there is a very strong trend of the mtGenome SNPs further resolving large groups based on mtCR SNPs, the single case in haplogroup D demonstrates why ideally one should sequence the entire mtGenome including the mtCR to use this genome to its complete capacity. Using AMOVA, the current dataset shows that there is little need to be concerned with whether a dog is classified as purebred or mixed or knowing the geographic location within the continental United States from which a sample was obtained. We do see evidence that it is necessary to collect multiple individuals of the same breed for a comprehensive mitochondrial SNP database. This is the first study to report SNP variation outside of the mtCR for the domestic dog. Our data demonstrate the usefulness of the entire mtGenome for forensic use in identifying domestic dog samples.

Acknowledgments

A portion of K. M. Webb's support came from a Selective Excellence Grant from The George Washington University. This work was supported by the National Institute of Justice through Grant 2004-DN-BX-K025 to M. W. Allard.

We thank Sheri Church, Gustavo Hormiga, Diana Johnson, Mark Wilson and two anonymous reviewers for careful review of this manuscript. The research was conducted at The George Washington University. The 698 new domestic dog samples were collected through donations from veterinary practices and private donors. We would like to thank Adobe Animal Hospital, Austin Vet Hospital, Caring Hands Animal Hospital, Del Paso Vet Clinic, Little River Vet Clinic, The College of Veterinary Medicine at Mississippi State University, Pet Medical Center of Las Vegas, Seneca Hill Animal Hospital, The Animal Clinic of Clifton, West Flamingo Animal Hospital for blood and tissue samples and 80 private donors who provided buccal swabs to add to our collection. Opinions or points of view expressed in this manuscript represent a consensus of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice or anyone who was acknowledged herein. The products and manufacturers discussed in this document are presented for informational purposes only and do not constitute product approval or enforcement by the U.S. Department of Justice.

References

1. Vigilant L. An evaluation of techniques for the extraction and amplification of DNA from naturally shed hairs. Biol Chem 1999;380(11):1329–31.
2. Bogenhagen D, Clayton D. The number of mitochondrial deoxyribonucleic acide genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial dexoyribonucleic acid. Journal of Biol Chem 1974;249:7991–5.
3. Nass M. Mitochondrial DNA. I. Intramitochondrial distribution and structural relations of single- and double-length circular DNA. J Mol Biol 1969;42:521–8.
4. Parsons TJ, Coble MD. Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. Croat Med J 2001;42(3):304–9.
5. Angleby H, Savolainen P. Forensic informativity of domestic dog mtDNA control region sequences. Forensic Sci Int 2005;154(2–3):99–110.
6. Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, et al. Mitochondrial DNA regions HVI and HVII population data. Forensic Sci Int 1999;103(1):23–35.
7. Gundry RL, Allard MW, Moretti TR, Honeycutt RL, Wilson MR, Monson KL, et al. Mitochondrial DNA analysis of the domestic dog: control region variation within and among breeds. J Forensic Sci 2007;52(3):562–72.
8. Okumura N, Ishiguro N, Nakano M, Matsui A, Sahara M. Intra- and interbreed genetic variations of mitochondrial DNA major non-coding regions in Japanese native dog breeds (Canis familiaris). Anim Genet 1996;27(6):397–405.
9. Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, et al. A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 1997;15(4):363–8.
10. Savolainen P, Lundeberg J. Forensic evidence based on mtDNA from dog and wolf hairs. J Forensic Sci 1999;44(1):77–81.
11. 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(4):593–600.
12. Webb KM, Allard MW. Identification of forensically informative SNPs in the domestic dog mitochondrial control region. J Forensic Sci 2009;54(2):doi:10.1111/j.1556-4029.2008.00953.x.
13. Pesole G, Gissi C, De Chirico A, Saccone C. Nucleotide substitution rate of mammalian mitochondrial genomes. J Mol Evol 1999;48(4):427–34.
14. Halverson J, Basten C. A PCR multiplex and database for forensic DNA identification of dogs. J Forensic Sci 2005;50(2):352–63.
15. Pereira L, Van Asch B, Amorim A. Standardisation of nomenclature for dog mtDNA D-loop: a prerequisite for launching a Canis familiaris database. Forensic Sci Int 2004;141(2–3):99–108.
16. Wetton JH, Higgs JE, Spriggs AC, Roney CA, Tsang CS, Foster AP. Mitochondrial profiling of dog hairs. Forensic Sci Int 2003;133(3):235–41.
17. Bjornerfeldt S, Webster MT, Vila C. Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Res 2006;16(8):990–4.
18. Kim KS, Lee SE, Jeong HW, Ha JH. The complete nucleotide sequence of the domestic dog (Canis familiaris) mitochondrial genome. Mol Phylogenet Evol 1998;10(2):210–20.
19. Woischnik M, Moraes CT. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res 2002;12(6):885–93.

20. Ishiguro N, Nakajima A, Horiuchi M, Shinagawa M. Multiple nuclear pseudogenes of mitochondrial DNA exist in the canine genome. Mamm Genome 2002;13(7):365–72.
21. Wilson MR, Allard MW, Monson K, Miller KW, Budowle B. Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region. Forensic Sci Int 2002;129(1):35–42.
22. Schneider S, Roessli D, Excoffier L. Arlequin: a software for population genetics data analysis, 2.0000 edn. Geneva: University of Geneva, Genetics and Biometry Lab, Dept. of Anthropology, 2000.
23. Nixon KC. Winclada, ver. 1.00.08. Ithaca, NY: Published by author, 2002. Available at http://www.cladistics.org.
24. Goloboff PA. NONA (NO NAME), 2nd edn. Tucuman, Argentina: Published by author, 1999.
25. Goloboff PA. Methods for faster parsimony analysis. Cladistics 1996; 12(3):199–220.
26. Zwickl DJ. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation. Austin, TX: The University of Texas at Austin, 2006.
27. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 1993;10(3):512–26.
28. Himmelberger AL, Spear TF, Satkoski JA, George DA, Garnica WT, Malladi VS, et al. Forensic utility of the mitochondrial hypervariable region 1 of domestic dogs, in conjunction with breed and geographic information. J Forensic Sci 2008;53(1):81–9.

Additional information and reprint requests:
Marc W. Allard, Ph.D.,
Food and Drug Administration
Office of Regulatory Science
Division of Microbiology
HFS-712
5100 Paint Branch Parkway
College Park, MD 20740
E-mail: mwallard@gwu.edu