ZooKeys 365: 381–411 (2013)
doi: 10.3897/zookeys.365.5859
www.zookeys.org

REVIEW ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Reviewing population studies for forensic purposes: Dog mitochondrial DNA

Sophie Verscheure[1,2], Thierry Backeljau[2,3], Stijn Desmyter[1]

1 *National Institute of Criminalistics and Criminology, Vilvoordsesteenweg 100, B-1120, Brussels, Belgium*
2 *University of Antwerp (Evolutionary Ecology Group), Groenenborgerlaan 171, B-2020, Antwerp, Belgium*
3 *Royal Belgian Institute of Natural Sciences (OD "Taxonomy and Phylogeny" and JEMU), Vautierstraat 29, B-1000, Brussels, Belgium*

Corresponding author: *Sophie Verscheure* (sophie.verscheure@just.fgov.be)

Academic editor: *M. De Meyer* | Received 25 June 2013 | Accepted 14 December 2013 | Published 30 December 2013

Citation: Verscheure S, Backeljau T, Desmyter S (2013) Reviewing population studies for forensic purposes: Dog mitochondrial DNA. In: Nagy ZT, Backeljau T, De Meyer M, Jordaens K (Eds) DNA barcoding: a practical tool for fundamental and applied biodiversity research. ZooKeys 365: 381–411. doi: 10.3897/zookeys.365.5859

Abstract

The identification of dog hair through mtDNA analysis has become increasingly important in the last 15 years, as it can provide associative evidence connecting victims and suspects. The evidential value of an mtDNA match between dog hair and its potential donor is determined by the random match probability of the haplotype. This probability is based on the haplotype's population frequency estimate. Consequently, implementing a population study representative of the population relevant to the forensic case is vital to the correct evaluation of the evidence. This paper reviews numerous published dog mtDNA studies and shows that many of these studies vary widely in sampling strategies and data quality. Therefore, several features influencing the representativeness of a population sample are discussed. Moreover, recommendations are provided on how to set up a dog mtDNA population study and how to decide whether or not to include published data. This review emphasizes the need for improved dog mtDNA population data for forensic purposes, including targeting the entire mitochondrial genome. In particular, the creation of a publicly available database of qualitative dog mtDNA population studies would improve the genetic analysis of dog traces in forensic casework.

Keywords

Forensics, Mitochondrial DNA, Dog, Random match probability, Population study, Sampling strategy

Introduction

Dogs (*Canis lupus familiaris*) are common and widespread in human society and hence, dog trace material is frequently encountered in forensic casework. Usually, this trace material involves hair, which is easily dispersed either through immediate contact with a dog or indirectly via an intermediate carrier, thus leaving a signature of the dog. Consequently, determining whether a particular dog could have donated the hair found at a crime scene may provide associative evidence (dis)connecting victims and suspects. For example, dog hairs could have been transferred from a victim's clothes to the trunk of a perpetrator's car during transportation of a body. Linking these hairs to the victim's dog could connect the suspect to the crime.

Most dog hairs collected at crime scenes are naturally shed and are in the telogen phase. As such, because they contain only limited amounts of, usually degraded, nuclear DNA (nDNA), they are ill suited for nDNA analysis. Conversely, mainly as a result of its high copy number and much smaller size (Nass 1969, Bogenhagen and Clayton 1974), mitochondrial DNA (mtDNA) is quantitatively and qualitatively better preserved than nDNA in telogenic hairs and hence is far more suitable for analysis, as e.g. demonstrated in Gagneux et al. (1997) and Allen et al. (1998). To identify the mammal taxon that shed the hair, DNA barcoding can be applied through analysis of an mtDNA marker with little variation within and sufficient variation among taxa, often a part of cytochrome *b* or cytochrome *c* oxidase subunit I in forensics (Linacre and Tobe 2011). On the other hand, in order to individualize dog hairs as accurately as possible, it is necessary to analyze mtDNA regions that show high variability among dogs and low intra-individual variation (heteroplasmy). As for human traces, this type of analysis focuses on the non-coding control region or D-loop (Wilson et al. 1993, Holland and Parsons 1999), which in dog mtDNA comprises about 1200 bp consisting of two hypervariable regions (HV-I and HV-II) separated by a Variable Number of Tandem Repeats (VNTR) region (Figure 1). This VNTR is a 10 bp tandem repeat with variable repeat numbers, both between and within individuals (length heteroplasmy). Because of its high level of length heteroplasmy, this repeat region is mostly not considered in forensics (Fridez et al. 1999). Several publications illustrate forensic casework involving control region analysis of dog traces, such as Savolainen and Lundeberg (1999), Schneider et al. (1999), Branicki et al. (2002), Aaspõllu and Kelve (2003), Halverson and Basten (2005) and Scharnhorst and Kanthaswamy (2011).

In general, mtDNA is maternally inherited (Sato and Sato 2013). In theory, this means that all dogs sharing a maternal line have the same mtDNA haplotype barring mutations. Hence, a match between the mtDNA of the dog hair found at a crime scene and that of a dog suspected of donating the trace, may be due to either of three possibilities: (1) the dog hair from the crime scene is from the suspected donor, (2) the hair from the crime scene is from a dog of the same maternal lineage as the suspected donor, (3) the mtDNA from the crime scene is by coincidence identical to that of the suspected donor. In order to assess the evidential weight of a match under the last scenario, one must calculate the haplotype's random match probability, the probability

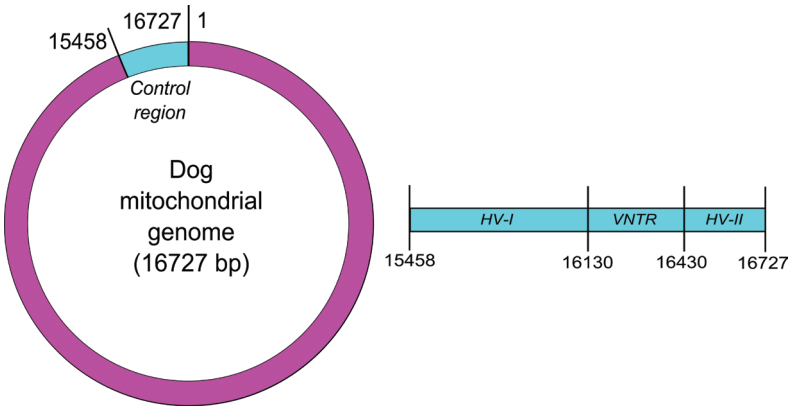


Figure 1. Position of the control region and its subregions within the Kim et al. (1998) reference dog mitochondrial genome.

that within a given population two randomly selected dogs will share the same haplotype by chance (Holland and Parsons 1999).

The random match probability is determined by the frequency estimate of the haplotype in the population of interest. The more common a haplotype, the higher is the probability that two dogs share this haplotype by chance, thus decreasing the evidential value of a match with this mtDNA type. Consequently, this sort of forensic applications requires the accurate estimation of haplotype frequencies in a population relevant to the criminal case.

The goal of this publication is to draw people's attention to the importance of implementing a dog mtDNA population study representative of the population of interest in a forensic case. It will provide an overview of the most important issues to keep in mind both when performing a population study of your own, as well as when considering to use published mtDNA data. First of all, sampling strategy characteristics are discussed such as sample size, maternal relatedness, breed status of the sampled dogs, and their geographic origin. Next, the importance of the quality of the sequence data is emphasized. In addition, the need to expand the sequenced DNA fragment in dog mtDNA studies is illustrated. Finally, the advantages of, and the criteria for, the assembly of an international, publicly available dog mtDNA population database of the highest quality, are pinpointed.

Estimating population frequencies of dog mtDNA haplotypes for forensic purposes

Background

The accuracy of haplotype frequency estimates almost entirely depends on the characteristics of the population sample that is used to represent the relevant population, i.e.

the population to which the donor of the trace is supposed to belong. Hence, biased population samples may lead to haplotype frequency estimates that diverge from the true population values.

To explore the impact of biased reference population samples in dog studies, we relied on current practices in human mtDNA population analyses and data derived from a selection of papers on haplotype variation in the dog mtDNA control region or the entire mitochondrial genome (mtGenome). Table 1 summarizes the main characteristics of the 58 dog studies used in this review. It includes studies with forensic aims, but also phylogenetic population and breed studies.

Dog mtDNA studies quite often do not meet the standards required for generating and publishing forensic human mtDNA population data. Briefly, these standards include: (1) providing a good documentation of the sampling strategy and a detailed description of the sampled individuals and the population, (2) avoiding sampling bias due to population substructure, (3) applying high quality mtDNA sequencing protocols and describing them clearly, (4) avoiding errors by handling and transferring data electronically, (5) performing quality checks of the generated data by e.g. haplogrouping or quasi-median network analysis and (6) making the full sequences publicly and electronically available preferably through either GenBank (Benson et al. 2013) or a forensic database such as EMPOP (Parson and Bandelt 2007, Parson and Dür 2007, Carracedo et al. 2010, Parson and Roewer 2010, Carracedo et al. 2013). These standards will be discussed here in relation to dog mtDNA studies.

Sampling strategy and its reporting

Strategies to sample mtDNA from dog populations are rarely well documented. Hence, it is often not clear to what extent the population samples adequately represent the populations from which they were drawn. Not seldom, sampling efforts are indeed limited to "sampling by convenience", i.e. relying on opportunistic sampling from locations as veterinary clinics and laboratories, dog shows, training schools and animal shelters. Obviously, it can be doubted whether these sampling locations are representative random samples of the "free" living relevant dog community (Parson and Bandelt 2007). Moreover, the types of sampling locations are often not even specified. In addition, basic information on the sampled dogs is often rudimentary, as many studies do not mention (1) how many dogs are mixed-breed or purebred, (2) to which breeds the dogs belong and/or (3) whether the geographic information provided refers to the region of origin of the breeds or to the actual region where the dogs were sampled (Table 1).

Several publications have provided recommendations on population sampling strategies for both dog and human mtDNA in forensics (Parson and Bandelt 2007, Pereira et al. 2010, Webb and Allard 2010, Linacre et al. 2011, Scharnhorst and Kanthaswamy 2011). Although more detailed guidelines are lacking, the main issue with

Table 1. Overview of the characteristics of sampling and sequence analysis in 58 canine mtDNA studies. Number of dogs sampled and, when specified in the publication, the number of dog breeds and mixed-breed, feral or village dogs in the sample; Origin of sample: new or extracted from previous studies as a comparison or to supplement the population sample (see reference numbers, except unpublished data by van Ash et al. (59); Koop et al. (60) and Shahid et al. (61)); Sampling region (or the geographic region of origin of included dog breeds if unclear from publication); Intention to avoid the inclusion of maternal relatives; GenBank accession numbers of new data are stated when applicable; un, unknown; s, skeletal remains of various age; ±, when variable, all sequences extracted from GenBank or the publication have this region in common; <, selected from this number of dogs from the same publication; * Larger region is mentioned in the publication, but only this part is available; Characteristics can differ from what is stated in publication if potential clerical errors were adapted, e.g. **publication states 246 instead of 233 as the sequences from reference 22 were included twice.

Publication			Sample						Sequence analysis	
	Reference	Aim of study	# Dogs	# Breeds	# Mixed-breed	Origin of sample	Sampling region	Avoid relatives	mtDNA region	Availability of new sequence data
1	Rothuizen et al. (1995)	mtDNA variability study	11	11	0	new	The Netherlands	un	Repeat region*	Publication
2	Okumura et al. (1996)	Phylogenetic breed study	94	24	0	new	Japan	un	15458-16129 16420-16727	D83599-D83613 D83616-D83638
3	Savolainen et al. (1997)	Forensic, Phylogenetic Population study	102	52	0	new	Sweden	YES	15431-15687	Publication
4	Tsuda et al. (1997)	Phylogenetic population study	34	24	0	new	Japan, Korea, Mongolia, Indonesia	un	15458-16130*	AB007380-AB007403
5	Vilà et al. (1997)	Phylogenetic population study	140 15	67 un	5 un	new < 140 new	un	un	15431-15687 ± 15393-16076 ± 16508-16549	AF005280-AF005295 AF008143-AF008157 AF008168-AF008182
6	Kim et al. (1998)	First complete dog mtGenome	1	1	0	new	Korea	YES	1-16727	U96639
7	Okumura et al. (1999)	Phylogenetic breed study	74s 84 94	un un 24	un un 0	new new 2	Japan	un	15483-15679 15458-16129 16420-16727	AB031089-AB031107
8	Vilà et al. (1999)	Phylogenetic breed study	19 140	1 67	0 5	new 5	US, Mexico un	YES	15431-15687* 15431-15687	Publication
9	Randi et al. (2000)	Phylogenetic population study	41 9	30 0	0 9	new new	Switzerland Italy	un	15458-16000	AF115704-AF115718
10	Kim et al. (2001)	Phylogenetic breed study	25	11	0	new	Korea	NO	15622-16030	AF064569-AF064579 AF064581-AF064585
11	Branicki et al. (2002)	Forensic population study	12	11	1	new	Poland	un	15431-15687	AF345977-AF345982

Publication			Sample						Sequence analysis	
	Reference	Aim of study	# Dogs	# Breeds	# Mixed-breed	Origin of sample	Sampling region	Avoid relatives	mtDNA region	Availability of new sequence data
12	Savolainen et al. (2002)	Phylogenetic population study	526 128	un	un	new 2, 4	Europe, Asia, Africa, Arctic America	un	± 15458-16039	AF531654-AF531741
13	Takahasi et al. (2002)	Inheritable disorder study	365	49	un	new	Japan	un	15458-16055	AB055010-AB055055
14	Valière et al. (2003)	Phylogenetic population study	50 15	un un	un un	new new	France, Switzerland France, Portugal	un	± 15519-15746	AF487730-AF487735 (excl. AF487732) AF487747-AF487751 AF338772-AF338788
15	Wetton et al. (2003)	Forensic population study	105 246	un	un	new 2, 3, 9	UK Japan, Switzerland, Italy, Sweden	un	15431-16030 ± 15458-15687	AY928903-AY928932
16	Pereira et al. (2004)	Catalogue of published datasets	58 1089	1 un	0 un	59 2, 4, 10, 12, 13, 15	Portugal Europe, Asia, Africa, Arctic America	un	15458-16039 ± 15622-16030	Publication
17	Savolainen et al. (2004)	Phylogenetic population study	22 19s 654	un un un	un un un	new new 2, 4, 12	SE-Asia, India Polynesia Europe, Asia, Africa, Arctic America	un	15458-16039 15458-15720 ± 15458-16039	AY660647-AY660650 Publication
18	Sharma et al. (2004)	Phylogenetic population study	24	0	24	new	India	un	15443-15783	AY333727-AY333737
19	Angleby and Savolainen (2005)	Forensic & Phylogenetic Population study	35 74 758	19 52 un	9 2 un	new new 2, 4, 12, 15	Germany Europe Europe, Asia, Africa, Arctic America	YES	15458-16039 ± 15458-16030	AY656703-AY656710
20	Halverson and Basten (2005)	Forensic population study	348	88	45	new	US	un	15431-16085	Not published
21	van Asch et al. (2005)	Phylogenetic breed study	143 144	4 9	0 0	new 2, 4, 12, 13	Portugal Europe, Asia, Africa, Arctic America	YES	15372-16083 ± 15458-16030	Publication
22	Björnerfeldt et al. (2006)	Phylogenetic population study	88 14	53 13	0 0	new < 88 new	Sweden	un	part of HV-I 1-16727	Not published DQ480489-DQ480502

Publication			Sample						Sequence analysis	
	Reference	Aim of study	# Dogs	# Breeds	# Mixed-breed	Origin of sample	Sampling region	Avoid relatives	mtDNA region	Availability of new sequence data
23	Pires et al. (2006)	Phylogenetic breed study	143 21	11 0	0 21	new new	Portugal, Spain, Morocco Portugal, Azores, Tunisia	YES	15211-16096	AY706476-AY706524
24	Ryabinina (2006)	Phylogenetic breed study	84 20	3 2	0 0	new 12	Russia Turkey	un	15458-15778 ± 15458-16039	DQ403817-DQ403837
25	Sundqvist et al. (2006)	Phylogenetic breed study	100	20	0	new	Sweden	un	15431-15687	Publication
26	Eichmann and Parson (2007)	Forensic population study	133	46	38	new	Austria	un	15458-16727	Publication
27	Gundry et al. (2007)	Forensic population study Forensic breed study	61 64	41 2	0 0	new new	US	un	15455-16727	AY240030-AY240157 (excluding AY240073 AY240094, AY240155)
28	Baute et al. (2008)	Forensic population study	83 159	30	0	new 27, 30	US	un	15595-15654	Publication
29	Hassell et al. (2008)	Forensic population study Forensic breed study	96 15	79 1	0 0	new new	UK	un	15458-16039 15458-16131 16428-16727	Not published
30	Himmelberger et al. (2008)	Forensic population study	36 22 179	11 un	20 un	new 60 2, 4, 5, 6, 10, 27	US (California) un Europe, Asia, North-America	un	15456-16063 15433-16139 ± 15622-16030	EF122413-EF122428 AF098126-AF098147
31	Parra et al. (2008)	Phylogenetic breed study	52	5	0	new	Spain	un	15458-16105	EF380216-EF380225
32	Baranowska et al. (2009)	Inheritable disorder study	7	1	0	new	Sweden	NO	1-16727	FJ817358-FJ817364
33	Boyko et al. (2009)	Phylogenetic population study	309 17 un	0 0 un	309 17 un	new new 12, 23	Egypt, Uganda, Namibia US (mostly Puerto Rico) East-Asia, Africa	YES	± 15454-16075 ± 15458-16039	GQ375164-GQ375213
34	Desmyter and Comblez (2009)	Forensic population study	117	60	24	new	Belgium	YES	15458-16130 16431-16727	Not published
35	Koban et al. (2009)	Phylogenetic breed study	114 un	2 un	0 un	new 12	Turkey Europe, Asia, Africa	YES	15458-16039	EF660078-EF660191

Publication			Sample						Sequence analysis	
	Reference	Aim of study	# Dogs	# Breeds	# Mixed-breed	Origin of sample	Sampling region	Avoid relatives	mtDNA region	Availability of new sequence data
36	Pang et al. (2009)	Phylogenetic population study	907	un	un	new, 61	Old World, Arctic America	un	± 15458-16039	EU816456-EU816557
			669	un	un	2, 4, 6, 12, 22			1-15511	
			135	un	un	< 907 + 669			15535-16039 16551-16727	EU789638-EU789786
			34	un	un	6, 22, 61			1-16727	AY656737-AY656755
37	Webb and Allard 2009a	Forensic population study	427	139	118	new	US	YES	± 15458-16114 ± 16484-16727	EU223385-EU223811
			125			27			15455-16727	
38	Webb and Allard 2009b	Forensic population study	64	43	11	37	US	YES	± 1-16129 ± 16434-16727	EU408245-EU408308
			15	14	0	6, 22	Korea, Sweden		1-16727	
39	Muñoz-Fuentes et al. (2010)	Phylogenetic population study	29	un	un	new	Canada	un	15361-15785	FN298190-FN298218
40	Smalling et al. (2010)	Forensic population study	220	0	220	new	US	YES	15456-16063	FJ501174-FJ501203
			429			30, 37			± 15458-16063	
41	Ardalan et al. (2011)	Phylogenetic population study	325	un	un	new	Europe, SW-Asia	un	15458-16039	HQ261489 HQ452418- HQ452423
			1576	un	un	2, 4, 6, 12, 22, 36, 61	Old World, Arctic America		± 15458-16039	HQ452432- HQ452433 HQ452466- HQ452477
42	Brown et al. (2011)	Phylogenetic population study	200	0	200	new	Middle East/SW-Asia	un	15482-15867	HQ287728- HQ287744
			231	0	231	new	SE-Asia			
			1576	un	un	2, 4, 6, 12, 22, 36, 61	Old World, Arctic America		± 15458-16039	
43	Castroviejo-Fisher et al. (2011)	Phylogenetic population study	371	0	371	new	the Americas	un	± 15491-15755	HQ126702- HQ127072
			29	un	un	39				
44	Klütsch et al. 2011a	Phylogenetic population study	280	33	0	new	Europe, Arctic America, East-Asia	YES	15458-16039	GQ896338-GQ896345
			234			36			± 15458-16039	

	Publication		Sample						Sequence analysis	
	Reference	Aim of study	# Dogs	# Breeds	# Mixed-breed	Origin of sample	Sampling region	Avoid relatives	mtDNA region	Availability of new sequence data
45	Klütsch et al. 2011b	Point heteroplasmy pedigree study	180 131	18 2	0 0	new new	Europe, Arctic America, East-Asia	NO	15458-16039	Publication
46	Kropatsch et al. (2011)	Phylogenetic breed study	77 34	26 1	0 0	new new	Germany	NO	15458-16124	Publication
47	Li et al. (2011)	Phylogenetic breed study	1 33	1 un	0 un	new 22, 32, 38, 61	China Sweden, US	YES	1-16727	HM048871
48	Sindičić et al. (2011)	Forensic species ID & Phylogenetic population study	20	0	20	new	Croatia	un	15465-15744	GU324475-GU324486
49	Bekaert et al. (2012)	Validation of forensic analysis method	41 550	29	3	new 27, 37	Belgium US	un	± 15458-16092 ± 16474-16703 ± 15458-16114 ± 16484-16727	HM561524-HM561546 HQ845266-HQ845282
50	Chakirou et al. (2012)	Phylogenetic breed study	78	3	0	new	Romania	YES	± 15251-16068	HE687017-HE687019
51	Desmyter and Gijsbers (2012)	Forensic population study Forensic breed study	208 778 107 337	60 6 6	68 0 0	new, 34 15, 26, 27, 37 new < 208 new, 13, 19, 26, 27, 37	Belgium UK, Austria, US Belgium Worldwide	YES	15458-16129 16430-16727 ± 15458-16030 15458-16129 16430-16727 ± 15458-16039	HM560872-HM560932
52 53	Głażewska et al. (2012) & Głażewska and Prusak (2012)	Phylogenetic breed study	34 un	2 un	0 un	new GenBank	Poland Worldwide	NO	15426-16085	HM007196-HM007200
54	Imes et al. (2012)	Forensic population study	100 233**	98 un	0 un	new 6, 22, 36, 38, 61	US, Australia, Canada, Columbia, Uruguay Worldwide	YES	± 1-16129 ± 16430-16727 ± 1-15511 ± 15535-16039 ± 16551-16727	JF342807-JF342906
55	Li and Zhang (2012)	Phylogenetic breed study	47 439	1 un	0 un	new GenBank	Tibet, surrounding areas Worldwide	YES	± 582 bp of control region	Not published

Publication			Sample						Sequence analysis	
	Reference	Aim of study	# Dogs	# Breeds	# Mixed-breed	Origin of sample	Sampling region	Avoid relatives	mtDNA region	Availability of new sequence data
56	Oskarsson et al. (2012)	Phylogenetic population study	305	un	un	new	SE-Asia, E-Asia	YES	15458-16039	HQ452439-HQ452465
			350	un	un	4, 12, 36				
			1224	un	un	2, 4, 6, 12, 22, 36, 61	Old World, Arctic America		± 15458-16039	
			19s	un	un	17	Polynesia		15458-15720	
57	Brown et al. (2013)	Phylogenetic population study	20s	un	un	new	Alaska, Greenland	un	± 367 bp of HV-I	JX185397
			51	1	0	new	Arctic America		± 15580-16016	
			78	2	0	2, 12, 36, 44			± 15458-16039	
58	Suárez et al. (2013)	Phylogenetic breed study	324	5	0	new	Canary Islands	YES	15361-16086	Publication
			986	un	un	15, 26, 27, 34, 37, 51	UK, Austria, Belgium, US		± 15458-16030	

sampling a population in a representative manner is to avoid over- and underestimating haplotype frequencies. Sampling bias causes regarding the number and features of sampled individuals will be discussed in relation to dog mtDNA.

Sample size

Using a random subsampling method (Pereira et al. 2004a), Webb and Allard (2010) assessed the influence of increased sample size on the distribution of haplotype frequency estimates in dog mtDNA population samples. They predicted that adding another 100 dogs to sample sizes of less than 650 dogs for HV-I and 750 dogs for HV-I and -II increases estimates of e.g. haplotype number and exclusion probability (i.e. the probability that two randomly chosen dogs from a sample have different haplotypes) with ≥ 5%. Table 1 shows that unless data are pooled, the majority of forensic dog control region population studies have rather small sample sizes of about 100 or fewer dogs.

Generally, the number of observed haplotypes increases with sample size (Table 2), while the proportion of rare haplotypes (i.e. encountered only once or twice) goes down. Consequently, exclusion probability largely remains the same with sample size expansion (Webb and Allard 2010) (Table 2). Under-sampling the population particularly affects the frequencies of haplotypes that remain rare while increasing sample size (Holland and Parsons 1999). This overestimation of rare haplotypes is illustrated when comparing nine forensic dog mtDNA studies. Many of the haplotypes with the highest frequencies in population samples of ≤ 100 dogs, have lower frequencies in larger sized studies (Table 2). For example, haplotype C5 occurs in 4.9% of the 61 dogs in the Gundry et al. (2007) study, while its frequency estimate is maximum 0.5% in other US studies in Table 2. Limited sample size thus tends to overestimate haplotype frequencies, which decreases the evidential value of an mtDNA match. Since overestimations do not inflate the risk of incriminating a false suspect, under-sampling can be deemed a conservative error (Salas et al. 2007).

Maternal relationships

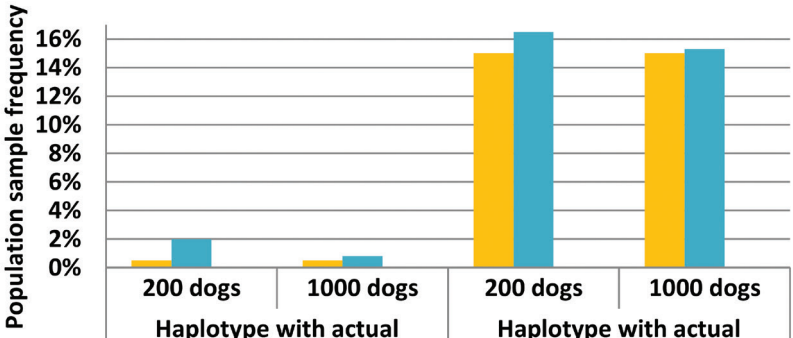
A randomized population sample for forensics should be allowed to include relatives if it is supposed to be unbiased (Brenner 2010). However, many population samples are assembled by convenience and could therefore contain more maternal relatives than expected from a randomized sample (Bodner et al. 2011).

The impact of a biased inclusion of maternal relatives in a forensic population study is rarely addressed, but generally decreases the genetic diversity of the population sample (Webb and Allard 2009a, Webb and Allard 2010). In small population samples, it particularly affects the risk of over-representing rare haplotypes (Bodner et al. 2011). By way of example, Figure 2 demonstrates that the impact of including 4 maternally related dogs is 5 times higher in a sample of 200 compared to 1000 dogs. In the smaller sample, the biased inclusion of 4 maternal relatives sharing a rare haplotype

Table 2. Comparison of haplotype number, P_E and haplotypes with the 10 highest frequencies in selected dog mtDNA studies. Exclusion probability (P_E) is based on the part of the control region studied in the publication (further details on exact region in Table 1) excluding the repeat region; characteristics can differ from publication if potential clerical errors were adapted; the 3 universally most frequent haplotypes are in bold (**A11, B1** and **A17**); (x) US and a minority from Australia, Canada, Uruguay and Columbia; (xx) Haplotype names are analogous to Savolainen et al. 2002, Savolainen et al. 2004, Angleby and Savolainen 2005, Pang et al. 2009, Klütsch et al. 201, Ardalan et al. 2011 and Oskarsson et al. 2012, or are in italic when unavailable and publication name was used; (xxx) Haplotypes are based on 15458-16039 except for Wetton et al. 2003 (15458-16030). Only the Kim et al. (1998) reference nucleotide was considered in case of a heteroplasmic site.

	Population studies for forensic purposes																		**Breed studies for forensic or phylogenetic purposes**					
	Europe								**US**															
	Wetton et al. 2003		**Angleby and Savolainen 2005**		**Eichmann and Parson 2007**		**Desmyter and Gijsbers 2012**		**Gundry et al. 2007**		**Himmelberger et al. 2008**		**Webb and Allard 2009a**		**Smalling et al. 2010**		**Imes et al. 2012**		**Okumura et al. 1996**		**Gundry et al. 2007**		**Suárez et al. 2013**	
Sampling region	UK		Germany		Austria		Belgium		US		US		US		US		US (x)		Japan		US		Canary Islands	
Studied part of control region (CR)	HV-I		HV-I		entire CR		HV-I+II		entire CR		HV-I		HV-I+II		HV-I		HV-I		HV-I+II		entire CR		HV-I	
# Dogs (# breeds/# mixed-breed)	105(un/un)		35(19/9)		133 (46/38)		208 (60/68)		61(41/0)		36(11/20)		552 (139/118)		649		100(98/0)		94(24/0)		64(2/0)		324(5/0)	
# Haplotypes	31		13		40		58		32		16		104		71		34		38		13		16	
Exclusion probability	0.93		0.86		0.93		0.92		0.93		0.89		0.96		0.92		0.91		0.93		0.8		0.86	
Haplotypes with 10 highest frequency estimates (%) (xx) (xxx)	**B1**	14.2	**A11**	28.6	**A11**	18.0	**B1**	16.8	**A11**	14.8	**B1**	16.7	**B1**	15.6	**B1**	18.0	**B1**	17.0	A18	13.8	A33	29.7	**B1**	25.9
	A18	11.4	**A17**	14.3	**A17**	12.0	**A11**	15.4	**B1**	13.1	**A11**	13.9	**A17**	10.9	**A11**	12.6	**A11**	15.0	A68	10.6	A16	28.1	**A17**	14.5
	A17	10.5	**B1**	14.3	**B1**	12.0	**A17**	15.4	A18	9.8	A16	13.9	**A11**	10.3	**A17**	11.7	A18	14.0	C3	10.6	**B1**	26.6	A20	14.2
	A2	8.6	C3	8.6	A19	8.3	A19	6.7	**A17**	6.6	**A17**	13.9	A18	9.2	A18	11.1	**A17**	8.0	**A17**	9.6	A5	6.3	B6	14.2
	A11	8.6	A2	5.7	A2	6.8	A18	5.8	C3	4.9	A18	11.1	A16	6.7	A16	6.2	A2	7.0	B14	8.5	*Gundry_24*	6.3	A19	8.3

	Population studies for forensic purposes																				Breed studies for forensic or phylogenetic purposes					
	Europe								US																	
	Wetton et al. 2003		Angleby and Savolainen 2005		Eichmann and Parson 2007		Desmyter and Gijsbers 2012		Gundry et al. 2007		Himmelberger et al. 2008		Webb and Allard 2009a		Smalling et al. 2010		Ines et al. 2012		Okumura et al. 1996		Gundry et al. 2007		Suárez et al. 2013	
Haplotypes with 10 highest frequency estimates (%) (xx) (xxx)	A19	5.7	A19	5.7	A18	6.8	A16	4.8	C5	4.9	A1	2.8	A33	3.4	A19	3.1	A22	4.0	A19	4.3	**A11**	1.6	**A11**	6.5
	A16	4.8	B6	5.7	A16	3.8	A22	3.8	A1	3.3	A15	2.8	C3	3.1	C3	3.1	A5	3.0	A2	3.2	**A17**	1.6	A22	4.9
	A20	3.8	A5	2.9	A1	3.0	A2	2.4	A2	3.3	A22	2.8	A2	2.5	A2	2.5	A16	3.0	**A11**	3.2			*A17+*	2.8
	A26	2.9	A20	2.9	A153	3.0	C1	2.4	A64	3.3	A26	2.8	A19	2.4	A5	1.7	*A167**	3.0	**B1**	3.2			A18	2.8
	A1	2.9	A33	2.9	A22	2.3	C2	2.4	C2	3.3	A28	2.8	A5	2.0	A22	1.7	A19	2.0	A29	2.1			A33	2.5
	C1	2.9	A44	2.9	A26	2.3			*Gundry_31*	3.3	A29	2.8			C2	1.7	A24	2.0	A70	2.1				
			A70	2.9	A33	2.3					A64	2.8							A72	2.1				
			A82	2.9	C1	2.3					A140	2.8							B12	2.1				
											A156	2.8							C1	2.1				
											B3	2.8												
											C3	2.8												



no maternal relatives	0.50%	0.50%	15.00%	15.00%
4 maternal relatives	2.00%	0.80%	16.50%	15.30%
Frequency increase	1.50%	0.30%	1.50%	0.30%

Figure 2. Impact of including maternally related dogs in population samples of 200 versus 1000 dogs on the estimation of the frequencies of rare haplotypes.

even causes its population sample frequency to be quadrupled. Moreover, even when observed only once in a population study, a haplotype that is rare in a population is already typically overrepresented in that sample (Holland and Parsons 1999). Therefore, oversampling maternal relatives should be avoided.

Although not specifically mentioned as a criterion for human mtDNA data in the international EMPOP database (Parson and Dür 2007), published population studies submitted to EMPOP do state that, as far as could be ascertained, the sampled individuals are unrelated. Examples are Brandstätter et al. (2007), Irwin et al. (2007), Saunier et al. (2009) and Prieto et al. (2011). Moreover, maternal relatives are removed in database updates. For human mtDNA population studies, it has been recommended to assess familial relationships by screening both available donor information and nDNA variation using microsatellites (Bodner et al. 2011).

For dog studies there is no consistent practice in dealing with maternal relationships in the population samples. Only about half of the 58 dog mtDNA studies in Table 1 mention whether or not they had the intention to avoid maternal relatives. Obviously, the usefulness of studies that do not provide this information may be doubtful in forensics. How maternal relationships were assessed is often not specified either, but it usually involves collecting information about the dogs from their owners. These background records can be used to verify whether dogs sharing a haplotype are e.g. from the same breed or whether their places of residence or those of their parents coincide (Webb and Allard 2009a, Desmyter and Gijsbers 2012). However, since dogs can have lots of offspring, there could be many maternal relatives, and these may be hard to track. Also, information provided by owners is not necessarily reliable (Webb and Allard 2009a) or available, and even registered pedigree records can be erroneous or incomplete (Kropatsch et al. 2011).

Purebred versus mixed-breed dogs

Another characteristic that can affect the haplotype frequency distribution in a population sample is potential population substructure due to the existence of dog breeds. Indeed, although generally mtDNA does not allow dogs to be grouped into their respective breeds (Okumura et al. 1996, Savolainen et al. 1997, Tsuda et al. 1997, Vilà et al. 1997, Kim et al. 2001, Wetton et al. 2003, Angleby and Savolainen 2005, van Asch et al. 2005, Pires et al. 2006, Sundqvist et al. 2006, Eichmann and Parson 2007, Gundry et al. 2007, Himmelberger et al. 2008, Parra et al. 2008, Desmyter and Comblez 2009, Kropatsch et al. 2011, Bekaert et al. 2012, Desmyter and Gijsbers 2012, Suárez et al. 2013), haplotype frequencies can differ between breeds, as well as between specific breeds and the entire dog mtDNA gene pool (Savolainen et al. 1997, Vilà et al. 1999, Angleby and Savolainen 2005, van Asch et al. 2005, Pires et al. 2006, Ryabinina 2006, Eichmann and Parson 2007, Gundry et al. 2007, Hassell et al. 2008, Himmelberger et al. 2008, Parra et al. 2008, Koban et al. 2009, Webb and Allard 2009a, Kropatsch et al. 2011, Desmyter and Gijsbers 2012, Brown et al. 2013, Suárez et al. 2013). Because of the overrepresentation of certain haplotypes in specific breeds in comparison to the general dog population, a dog trace mtDNA type might provide an indication about the breed(s) to which it may belong. However, such information should be used with caution and police investigations should not only focus on the more likely breed (Angleby and Savolainen 2005, Hassell et al. 2008, Desmyter and Gijsbers 2012).

Obviously, the over- and underrepresentation of particular breeds in a population sample compared to the population from which the sample is drawn, may bias haplotype frequency estimates (Desmyter and Gijsbers 2012). Therefore, in theory, dogs should be randomly sampled in order to correctly represent the breed composition of the population of interest (Scharnhorst and Kanthaswamy 2011). In addition, it is recommended that population samples reflect the actual proportions of mixed-breed versus purebred dogs in the population. However, mixed-breed dogs are often underrepresented in population studies (Smalling et al. 2010) and many population studies even only include purebred dogs (Table 1). Moreover, since most samples are collected by convenience, deviations from the actual breed composition of the population and overrepresentation of rare dog breeds are to be expected. This should be taken into account when using this sort of data in forensic casework (Eichmann and Parson 2007).

Against this background, Savolainen et al. (1997) attempted to adjust the number of dogs per breed in their population sample, so as to more accurately reflect the countrywide breed composition in Sweden. Later, Angleby and Savolainen (2005) stated that the exclusion probability of population samples containing dogs of the 20 most common breeds in Sweden represent the Swedish population more accurately than population samples containing the 100 most common breeds. Still, the number of dogs per breed and the overall number of breeds in a sample can be overestimated, since these data largely depend on the owner's subjective opinion (Himmelberger et al. 2008, Webb and Allard 2009a).

Some authors have indicated that population studies specific for single breeds may be forensically relevant in the rare event that the breed of the dog that donated the crime scene trace is known, for example by eye-witness reports (Savolainen et al. 1997, Wetton et al. 2003, Desmyter and Gijsbers 2012). Obviously, the evidential value of an mtDNA match can be quite different when based on a general rather than a breed-specific population study. Studies focused on specific breeds have been published, mostly aiming at verifying the accuracy of pedigree records and tracing its population genetic features (e.g. demographic history, region of origin, hybridization events, etc.). Examples are Vilà et al. (1999), van Asch et al. (2005), Kropatsch et al. (2011) and Suárez et al. (2013).

Including pedigree data can improve intra-breed mtDNA diversity studies. In theory, an appropriate selection of representative individuals from existing maternal lines from pedigrees allows to capture all mtDNA haplotypes of a breed within a population while minimizing the amount of laboratory work. The frequencies of these haplotypes can be estimated from the numbers of offspring in each maternal line in the breed population (Głażewska et al. 2013). Of course, to this end pedigree records need to be accurate and complete (Głażewska et al. 2013). Unfortunately, this is not always the case, as has been shown in e.g. Weimaraner dogs (Kropatsch et al. 2011).

Analyzing the haplotype frequency distribution within breeds can also give insight into differences between published population studies. An example of the impact of breed associated sample bias was given by Desmyter and Gijsbers (2012). These authors noted that the US population sample of Webb and Allard (2009a) included 64 dogs of two Retriever breeds from Gundry et al. (2007). This could have biased the frequency estimates of haplotypes A16 and A33 in the US sample, since these haplotypes are very common in Retrievers (Desmyter and Gijsbers 2012). Additionally, mtDNA studies focusing on specific breeds rather than on entire populations, clearly show lower amounts of variation (expressed in terms of exclusion probability) than population studies from similar geographical regions (Table 2). Also, the sets of haplotypes with the ten highest frequencies can be quite different (Table 2). In order to compensate for purebred related biases, Himmelberger et al. (2008) increased the number of mixed-breed dogs in their US population sample and claimed that in this way their sample was more representative than previous US dog mtDNA population samples. However, their sample still showed an unusually high frequency of haplotype A16 (Table 2), most probably because their sample was limited to only 36 dogs, 13 of which were either purebred or mixed-breed Retrievers.

Geographic origin

To evaluate the significance of a haplotype match between a dog trace and its suspected donor, a population sample should reliably reflect the population to which the donor of the trace is supposed to belong. As such, one might wonder about the importance of the geographic origin of the sampled dogs in a sampling strategy.

Probably the most important macrogeographic issue to consider in dog studies, is the fact that dog populations in Southeast Asia show almost the entire dog mtDNA diversity, while elsewhere in the world only parts of this diversity is present (Savolainen et al. 2002, Pang et al. 2009, Ardalan et al. 2011, Brown et al. 2011). This suggests that SE Asia is the region where dogs were first domesticated and from where domesticated dogs were spread throughout the rest of the world (Savolainen et al. 2002, Pang et al. 2009). Another noticeable macrogeographic structuring in dog mtDNA is that haplotype group d1 is almost exclusively found in Scandinavian and Finnish breeds, in which sometimes over 50% of the dogs have a d1 haplotype (Klütsch et al. 2011a). Obviously, this sort of macrogeographic mtDNA differentiation should be considered in population sampling, since oversampling dog breeds of SE Asian or Scandinavian/Finnish origin in local population samples elsewhere in the world can bias haplotype frequency estimates. For example, Angleby and Savolainen (2005) demonstrated that dogs of East Asian origin in Europe carried a number of haplotypes that are absent in native European breeds. Moreover, the frequencies of globally common haplotypes differ between Asian and European samples (Angleby and Savolainen 2005). This is also illustrated by the composition and frequency distribution of the most common dog haplotypes in the breed study from Japan by Okumura et al. (1996) and those in the forensic population studies from Europe and the US (Table 2).

Quality of nucleotide sequence data

The description of haplotypes is a source of error and confusion when comparing population studies. Typically, haplotypes are aligned to a reference sequence using software supplemented with annotation rules in order to record them unambiguously as an alpha-numeric code. This code is a shortened annotation of the sequence string, consisting of differences to the reference sequence. For example, the HV-I alpha-numeric code of haplotype A11 is 15639A, 15814T and 16025C (Angleby and Savolainen 2005). Analogous to human mtDNA analyses, Pereira et al. (2004b) recommended to set the L-strand of the first published complete dog mtGenome (Kim et al. 1998) as the reference standard. In order to identify different haplotypes and enable their comparison, Pereira et al. (2004b) listed a number of rules to align sequences to the Kim et al. reference (1998) and to unambiguously record polymorphisms. These rules are based on those for human mtDNA (Carracedo et al. 2000, Wilson et al. 2002b, Wilson et al. 2002b). Length heteroplasmy in the VNTR region of the dog's mtDNA control region complicates the numbering system of the nucleotide positions. To simplify this, Pereira et al. (2004b) decided that numbering the nucleotide positions after this repeat region should start at position 16430 regardless of the number of repeats (Figure 1). Nevertheless, even with a standard reference haplotype, a numbering system and annotation rules, variation can still be miscoded, such as for the polyC-polyT-polyC region from position 16661 to 16674 in HV-II (Table 3).

Table 3. Illustration of different annotations for the HV-II polyC-polyT-polyC haplotype with 6 C's, 8 T's and 2 C's. Annotation (1) was used by Gundry et al. (2007), while Eichmann and Parson (2007) and Desmyter and Gijsbers (2012) applied annotation (2) because of different alignments to the Kim et al. (1998) reference sequence of 3C8T3C.

#C#T#C	16661	16662	16663	16663.1	16663.2	16663.3	16664	16665	16666	16667	16668	16669	16670	16671	16672	16673	16674
3C8T3C	**C**	**C**	**C**	**-**	**-**	**-**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**C**	**C**	**C**
6C8T2C (1)	C	C	C	C	C	-	C	T	T	T	T	T	T	T	T	C	C
6C8T2C (2)	C	C	C	C	C	C	T	T	T	T	T	T	T	T	C	C	-

Haplotypes can also be denoted by names. However, it is not good practice to provide only haplotype names in publications, like e.g. Sundqvist et al. (2006) did. This introduces ambiguities if the same names are used elsewhere for other haplotypes. For the same reason, it is ill advised to use haplotype names that differ from GenBank entries, as was done by e.g. Smalling et al. (2010). The haplotype names established by Savolainen et al. (2002), Savolainen et al. (2004) and Angleby and Savolainen (2005), were expanded by Pang et al. (2009) and Webb and Allard (2009a), such that they both used the same names for different new haplotypes. Since then, names of new haplotypes often overlap between publications that are building further onto the names of both of these publications, e.g. Smalling et al. (2010), Ardalan et al. (2011), Klütsch et al. (2011a) and Imes et al. (2012). In addition, applying previously published haplotype names can be difficult because the analyzed mtDNA region may differ (Pereira et al. 2004b).

Mistakes occur relatively often while copying and editing sequence data. Therefore, guidelines have been published to minimize making these clerical errors and to detect them more easily (Bandelt et al. 2001, Bandelt et al. 2004, Yao et al. 2004, Salas et al. 2005). For example, alpha-numeric codes presented in the form of a matrix-based dot table are particularly error-prone and difficult to read (Parson and Bandelt 2007, Parson and Roewer 2010). In practice, several clerical errors have been observed in dog mtDNA studies. For example, alignment with the Kim et al. (1998) reference sequence of the GenBank entries corresponding to the HV-I codes in Table 2 of Imes et al. (2012), revealed several inconsistencies. A deletion at position 15932 in many haplotypes in Table 2 of Imes et al. (2012) cannot be observed in most of the GenBank entries. As such, this deletion defined two artificial haplotypes. Furthermore, Table 2 of Imes et al. (2012) did not include two haplotypes deposited in GenBank, while variant base 15665C was not recorded for haplotype A170*.

As more mtGenome data are generated, coding regions SNPs are encountered that appear to be characteristic for particular control region haplotypes and haplogroups (Verscheure, unpublished data). Such SNPs can help to indicate potential sequence or clerical errors. For example, the control region sequence of mtGenome haplotype A169* (A11 after removal of the deletion at 15932) belongs to haplogroup A (Imes et al. 2012), but the SNPs in the rest of its mtGenome are typical for haplogroup B. This might be due to artificial recombination, caused by mixing up amplicons from dif-

ferent individuals, either during laboratory work or data editing (Bandelt et al. 2001, Bandelt et al. 2004). Similarly, the entire mtGenome sequence of Imes et al. (2012) haplotype A167* is more typical of haplogroup C than of haplogroup A.

As shown above, deposition of sequence data in GenBank provides an opportunity to verify sequence data quality. Unfortunately, in contrast to good practice, 15 of the 58 studies reviewed here did not submit any sequence to GenBank, but only provided alpha-numeric codes or haplotype names. Moreover, several papers did not even disclose the haplotype sequences or their estimated population frequencies (Table 1). When studies did deposit sequences in GenBank, they did so either only for new haplotypes, for all observed haplotypes, or for all sampled dogs. These various practices may confound subsequent analyses. For example, Imes et al. (2012) extracted the mtGenomes of the same 14 dogs twice from GenBank, because these sequences were uploaded in GenBank both by Björnerfeldt et al. (2006) and Pang et al. (2009). Obviously, these duplicated datasets introduce bias in the estimation of mtGenome haplotype frequencies and mtDNA diversity in the Imes et al. (2012) study.

Dog mtDNA studies show a large variety of analysis methods as well. Consequently, the quality of these analyses might vary. Next to annotation issues, several sequence quality issues have been observed while reviewing dog mtDNA studies. For example, Webb and Allard (2009a) reported sequence reading difficulties in the HV-II region because of length heteroplasmy in the VNTR and the polyC-polyT-polyC region. Nevertheless, in about 190 sequences Webb and Allard (2009a) observed that positions 16430, 16431, 16432 and/or 16433 directly adjacent to the VNTR at the start of HV-II are deleted in comparison to the Kim et al. (1998) reference sequence. Since deletions have not been reported at these sites in any of the other reviewed studies, we suggest these should be considered missing data due to reading difficulties. Therefore, it is recommended to verify this issue using additional primers, other alignment software and visual inspection of the alignments. Webb and Allard (2009a) interpreted these sites as highly informative and counted sequences differing only at these positions (e.g. A1 and d) as different haplotypes. If these deletions indeed resulted from reading difficulties, then they artificially increased the haplotype number and exclusion probability. A second example is that about 65% of the mtGenome sequences deposited in GenBank by Imes et al. (2012) contain ambiguities outside the VNTR with up to 130 N's per sequence and stretches of up to 110 adjacent N's, i.e. ambiguous bases due to the presence of dye blobs (Imes et al. 2012). If such ambiguities occur at informative sites, then these sequence quality issues can affect the frequency estimates of mtGenome haplotypes and the SNPs that define them.

Thus, caution and proofreading is necessary for both new sequences and those extracted from papers and databases. Therefore, Berger et al. (2012) published a detailed workflow for generating high quality HV-I and -II data from dogs based on experience from human mtDNA analysis. For forensic mtDNA analysis, these and other authors recommend to sequence each position at least twice, preferably on both mtDNA strands, so as to minimize sequencing errors (Wilson et al. 1993, Carracedo et al. 2000, Tully et al. 2001, Parson and Bandelt 2007, Berger et al. 2012). Finally, when extracting

sequences from GenBank, it is important to realize that quality control of a database entry relies on the submitting scientist. Hence, it is not surprising that the reliability of GenBank data has been questioned, such as by Harris (2003) and Yao et al. (2009).

Exploring the entire mtGenome to improve discriminatory power

The majority of dogs have haplotypes that are frequent in most dog populations worldwide. As a result, even if there are many rare haplotypes, the discriminatory power of the dog mtDNA control region is limited (Savolainen et al. 1997, Wetton et al. 2003, Angleby and Savolainen 2005, Halverson and Basten 2005, Eichmann and Parson 2007, Gundry et al. 2007, Baute et al. 2008, Hassell et al. 2008, Himmelberger et al. 2008, Desmyter and Comblez 2009, Webb and Allard 2009a, Smalling et al. 2010, Desmyter and Gijsbers 2012, Imes et al. 2012). This is well illustrated by comparing the mtDNA characteristics of nine forensic population studies which all consider at least positions 15458 to 16030 in HV-I. Almost half of the sampled dogs have haplotypes B1, A11 or A17 with average population frequency estimates of 15.3%, 15.2% and 11.5%. In addition, many other frequent haplotypes are shared between samples (Table 2). Hence, dog mtDNA matches will often have limited forensic value.

Evidently, expanding the length of the surveyed sequence will increase the number of polymorphic sites and thus may improve the discriminatory power of the mtDNA control region in dogs. However, most population studies did not include HV-II and as such missed important variation that often allows splitting up HV-I haplotypes. Hence, sequencing at least both HV-I and HV-II is recommended for forensic population studies (Eichmann and Parson 2007, Gundry et al. 2007, Desmyter and Comblez 2009, Webb and Allard 2009a, Webb and Allard 2010, Desmyter and Gijsbers 2012, Imes et al. 2012).

A number of complete control region haplotypes still show high population frequencies. Therefore, it is advised to further increase the discriminatory power of dog mtDNA by surveying population samples for entire mtGenomes (Webb and Allard 2009a). This is indeed a trend in the last years with very promising results (Webb and Allard 2009b, Imes et al. 2012). However, the use of SNPs in the coding region in forensics will require many more mtGenome studies (Irwin et al. 2011).

Population study versus database

Not every forensic laboratory has the resources to conduct large-scale population studies. As such, supplementing smaller, local samples with published data allows capturing more mtDNA variability. However, this practice may bias the haplotype frequency distribution in the pooled sample compared to the population of interest, because of (1) sample heterogeneity, (2) inconsistent sequence quality, (3) clerical errors and (4)

the difficulty of sequence comparisons due to variation in sequence lengths, alignment procedures, and sequence annotation. Relying on a public dog mtDNA database instead of, or in addition to, published local population data may be a trustworthy alternative, provided that the sequences are carefully reviewed before inclusion in the database. As such, submitting population sample data to the database could be an obligatory quality check with which studies have to comply before they are published. This is often demanded for human mtDNA population data (Carracedo et al. 2010, Parson and Roewer 2010, Carracedo et al. 2013).

To establish a reliable dog mtDNA database, inspiration can be found in the European DNA profiling group (EDNAP) mtDNA population database (EMPOP) for human mtDNA haplotypes useful in forensic casework. EMPOP stresses the need for generating mtDNA sequence data of the highest quality (Parson et al. 2004, Parson and Dür 2007) and established guidelines to achieve this. Briefly, these guidelines recommend: (1) application of a high quality mtDNA determination method that covers the entire sequenced region at least twice, (2) electronic transfer and transcription of sequence results, (3) compliance to generally accepted alignment and annotation guidelines and (4) data verification through haplogrouping and quasi-median network analysis (Brandstätter et al. 2007, Parson and Dür 2007). Against this background, the interlaboratory study by van Asch et al. (2009) emphasized a similar need for such guidelines for dog mtDNA analyses.

Next to the need for high quality mtDNA population data from all around the world, three other important requirements for building a dog mtDNA database are discussed hereafter. Firstly, management by a central laboratory is indispensable to perform the quality assessment of submitted population samples, to maintain and update the database software and web portal, and to communicate about it to the users. After submission to EMPOP, this laboratory reviews the population sample data for errors by e.g. examining the raw sequence data and using quasi-median network analysis (Bandelt and Dür 2007, Parson and Dür 2007, Zimmermann et al. 2011). Indeed, allocating mtDNA sequences to specific haplogroups may indicate which mutations are expected and may help to detect potential artificial recombination (Bandelt et al. 2001, Bandelt et al. 2004, Bandelt et al. 2012). Additionally, EMPOP provides its users with software for network analysis that may point out potential errors within the data based on phylogenetic background information. Thorough phylogenetic knowledge of dog mtDNA haplotypes could allow the adaptation of such software for dog mtDNA population samples.

Secondly, the database should be searchable and provide tools for comparison of various mtDNA sequence ranges. EMPOP uses the SAM search engine, which translates the queried haplotype and all database entries into sequence strings that are more easily comparable than alpha-numeric codes. In this way, it avoids generating biased haplotype frequency estimates caused by alignment and annotation inconsistencies making that database entries remain undetected in a database search even if they are identical to the queried haplotype (Röck et al. 2011).

Finally, the database should sufficiently document background information on the specimens. This enables the selection of subsets of samples in the database relevant to a specific case, such as dogs from specific geographic regions, of particular breeds, etc. In casework, selection of a suitable dataset is vital to a correct evaluation of evidence. Weighing the evidence against several database subdivisions is recommended to consider which one provides the most appropriate and conservative estimate of a haplotype's random match probability (Salas et al. 2007).

FidoSearch™, a canine mtDNA database with search software, was developed for use in casework by the Institute of Pathology and Molecular Immunology in Porto, Portugal in collaboration with Mitotyping Technologies in Pennsylvania, USA (Melton et al. 2011). However, it is not publicly available and its data entries were assembled from GenBank. Hence, FidoSearch™ is not an appropriate alternative for the creation of a publicly available, high quality and comprehensive dog mtDNA database.

Conclusions

In order to meet forensic quality standards, a dog mtDNA population sample needs to be representative of the population of interest to the case. To this end, several recommendations can be made for performing and publishing a dog mtDNA population study for forensic purposes: (1) provide sufficiently detailed information on the population of interest, the sampling strategy and the sampled dogs, (2) include at least several hundred dogs in the population sample, (3) intend to avoid biased inclusion of maternal relatives, (4) use a population sample reflecting the dog population where the crime occurred, (5) the composition of the population sample in terms of purebred and mixed-breed dogs, groups of breeds of a particular geographic origin, and dogs belonging to specific breeds, should be proportional to the studied population, (6) apply a high quality and validated analytical methodology and run quality control steps to minimize the risk of errors during either laboratory work or data processing, (7) submit the haplotype sequence strings to a publicly available database such as GenBank and (8) follow the Pereira et al. (2004b) rules when converting haplotype sequences into alpha-numeric codes denoting differences in relation to the Kim et al. (1998) reference sequence. These recommendations also apply when supplementing your own data with published data. In addition, keep in mind that sequence files in a database such as GenBank do not provide raw sequence data and can hide ambiguous results.

All things considered, this review emphasizes the need for more forensically relevant, high quality dog mtDNA population studies. In addition, it stresses the need for a publicly available dog mtDNA population database that assembles easily comparable and thoroughly checked population data from all around the world. Finally, expanding mtDNA studies from the control region to the entire mtGenome is recommended to enhance the discriminatory power of forensic dog mtDNA analysis.

Acknowledgements

S. Verscheure is a PhD student at the University of Antwerp supported by a grant from the Belgian Federal Public Planning Service Science Policy. This work was conducted within the framework of FWO Research Community W0.009.11N "Belgian Network for DNA Barcoding".

References

Aaspóllu A, Kelve M (2003) The first criminal case in Estonia with dog's DNA data admitted as evidence. International Congress Series 1239: 847–851.

Allen M, Engström AS, Meyers S, Handt O, Saldeen T, von Haeseler A, Pääbo S, Gyllensten U (1998) Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. Journal of Forensic Sciences 43: 453–464.

Angleby H, Savolainen P (2005) Forensic informativity of domestic dog mtDNA control region sequences. Forensic Science International 154: 99–110. doi: 10.1016/j.forsciint.2004.09.132

Ardalan A, Klütsch CF, Zhang AB, Erdogan M, Uhlén M, Houshmand M, Tepeli C, Ashtiani SR, Savolainen P (2011) Comprehensive study of mtDNA among Southwest Asian dogs contradicts independent domestication of wolf, but implies dog-wolf hybridization. Ecology and Evolution 1: 373–385. doi: 10.1002/ece3.35

Bandelt HJ, Dür A (2007) Translating DNA data tables into quasi-median networks for parsimony analysis and error detection. Molecular Phylogenetics and Evolution 42: 256–271. doi: 10.1016/j.ympev.2006.07.013

Bandelt HJ, Lahermo P, Richards M, Macaulay V (2001) Detecting errors in mtDNA data by phylogenetic analysis. International Journal of Legal Medicine 115: 64–69. doi: 10.1007/s004140100228

Bandelt HJ, Salas A, Lutz-Bonengel S (2004) Artificial recombination in forensic mtDNA population databases. International Journal of Legal Medicine 118: 267–273. doi: 10.1007/s00414-004-0455-2

Bandelt HJ, van Oven M, Salas A (2012) Haplogrouping mitochondrial DNA sequences in Legal Medicine/Forensic Genetics. International Journal of Legal Medicine 126: 901–916. doi: 10.1007/s00414-012-0762-y

Baranowska I, Jäderlund KH, Nennesmo I, Holmqvist E, Heidrich N, Larsson NG, Andersson G, Wagner EG, Hedhammar Å, Wibom R, Andersson L (2009) Sensory ataxic neuropathy in golden retriever dogs is caused by a deletion in the mitochondrial tRNATyr gene. PLoS Genetics 5: e1000499. doi: 10.1371/journal.pgen.1000499

Baute DT, Satkoski JA, Spear TF, Smith DG, Dayton MR, Malladi VS, Goyal V, Kou A, Kinaga JL, Kanthaswamy S (2008) Analysis of forensic SNPs in the canine mtDNA HV1 mutational hotspot region. Journal of Forensic Sciences 53: 1325–1333. doi: 10.1111/j.1556-4029.2008.00880.x

Bekaert B, Larmuseau MH, Vanhove MP, Opdekamp A, Decorte R (2012) Automated DNA extraction of single dog hairs without roots for mitochondrial DNA analysis. Forensic Science International, Genetics 6: 277–281. doi: 10.1016/j.fsigen.2011.04.009

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. Nucleic Acids Research 41: D36–42. doi: 10.1093/nar/gks1195

Berger C, Berger B, Parson W (2012) Sequence analysis of the canine mitochondrial DNA control region from shed hair samples in criminal investigations. Methods in Molecular Biology 830: 331–348. doi: 10.1007/978-1-61779-461-2_23

Björnerfeldt S, Webster MT, Vilà C (2006) Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Research 16: 990–994. doi: 10.1101/gr.5117706

Bodner M, Irwin JA, Coble MD, Parson W (2011) Inspecting close maternal relatedness: Towards better mtDNA population samples in forensic databases. Forensic Science International, Genetics 5: 138–141. doi: 10.1016/j.fsigen.2010.10.001

Bogenhagen D, Clayton DA (1974) The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. Journal of Biological Chemistry 249: 7991–7995.

Boyko AR, Boyko RH, Boyko CM, Parker HG, Castelhano M, Corey L, Degenhardt JD, Auton A, Hedimbi M, Kityo R, Ostrander EA, Schoenebeck J, Todhunter RJ, Jones P, Bustamante CD (2009) Complex population structure in African village dogs and its implications for inferring dog domestication history. Proceedings of the National Academy of Sciences of the USA 106: 13903–13908. doi: 10.1073/pnas.0902129106

Brandstätter A, Niederstätter H, Pavlic M, Grubwieser P, Parson W (2007) Generating population data for the EMPOP database - an overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example. Forensic Science International 166: 164–175. doi: 10.1016/j.forsciint.2006.05.006

Branicki W, Kupiec T, Pawłowski R (2002) Analysis of dog mitochondrial DNA for forensic identification purposes. Problems of Forensic Sciences (Z Zagadnien Nauk Sadowych) 50 (L): 91–98.

Brenner CH (2010) Fundamental problem of forensic mathematics--the evidential value of a rare haplotype. Forensic Science International, Genetics 4: 281–291. doi: 10.1016/j.fsigen.2009.10.013

Brown SK, Darwent CM, Sacks BN (2013) Ancient DNA evidence for genetic continuity in arctic dogs. Journal of Archaeological Science 40: 1279–1288. doi: 10.1016/j.jas.2012.09.010

Brown SK, Pedersen NC, Jafarishorijeh S, Bannasch DL, Ahrens KD, Wu JT, Okon M, Sacks BN (2011) Phylogenetic distinctiveness of Middle Eastern and Southeast Asian village dog Y chromosomes illuminates dog origins. PLoS ONE 6: e28496. doi: 10.1371/journal.pone.0028496

Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Forensic Science International 110: 79–85. doi: 10.1016/S0379-0738(00)00161-4

Carracedo A, Butler JM, Gusmão L, Linacre A, Parson W, Roewer L, Schneider PM (2013) New guidelines for the publication of genetic population data. Forensic Science International, Genetics 7: 217–220. doi: 10.1016/j.fsigen.2013.01.001

Carracedo A, Butler JM, Gusmão L, Parson W, Roewer L, Schneider PM (2010) Publication of population data for forensic purposes. Forensic Science International, Genetics 4: 145–147. doi: 10.1016/j.fsigen.2010.02.001

Castroviejo-Fisher S, Skoglund P, Valadez R, Vilà C, Leonard JA (2011) Vanishing native American dog lineages. BMC Evolutionary Biology 11: 73. doi: 10.1186/1471-2148-11-73

Chakirou O, Vlaic A, Carşai TC, Bâlteanu VA, Coşier V (2012) Aspects regarding the molecular characterisation of the Romanian Shepherd dog breeds. Animal Biology & Animal Husbandry 4: 28–31.

Desmyter S, Comblez S (2009) Belgian dog mitochondrial DNA database for forensics. Forensic Science International, Genetics Supplement Series 2: 286–287. doi: 10.1016/j.fsigss.2009.08.110

Desmyter S, Gijsbers L (2012) Belgian canine population and purebred study for forensics by improved mitochondrial DNA sequencing. Forensic Science International, Genetics 6: 113–120. doi: 10.1016/j.fsigen.2011.03.011

Eichmann C, Parson W (2007) Molecular characterization of the canine mitochondrial DNA control region for forensic applications. International Journal of Legal Medicine 121: 411–416. doi: 10.1007/s00414-006-0143-5

Fridez F, Rochat S, Coquoz R (1999) Individual identification of cats and dogs using mitochondrial DNA tandem repeats? Science & Justice 39: 167–171. doi: 10.1016/S1355-0306(99)72042-3

Gagneux P, Boesch C, Woodruff DS (1997) Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. Molecular Ecology 6: 861–868. doi: 10.1111/j.1365-294X.1997.tb00140.x

Głażewska I, Prusak B (2012) Evaluation of the effectiveness of introducing new alleles into the gene pool of a rare dog breed: Polish Hound as the example. Czech Journal of Animal Science 57: 248–254.

Głażewska I, Prusak B, Gralak B (2013) Pedigrees as a source of information in mtDNA studies of dogs and horses. Animal Genetics 44: 227–230. doi: 10.1111/j.1365-2052.2012.02388.x

Głażewska I, Zielińska S, Prusak B (2012) Formation of a new dog population observed by pedigree and mtDNA analyses of the Polish Hovawart. Archiv Tierzucht 55: 391–401.

Gundry RL, Allard MW, Moretti TR, Honeycutt RL, Wilson MR, Monson KL, Foran DR (2007) Mitochondrial DNA analysis of the domestic dog: control region variation within and among breeds. Journal of Forensic Sciences 52: 562–572. doi: 10.1111/j.1556-4029.2007.00425.x

Halverson JL, Basten C (2005) Forensic DNA identification of animal-derived trace evidence: tools for linking victims and suspects. Croatian Medical Journal 46: 598–605.

Harris DJ (2003) Can you bank of GenBank. Trends in Ecology & Evolution 18: 317–319. doi: 10.1016/S0169-5347(03)00150-2

Hassell R, Heath P, Musgrave-Brown E, Ballard D, Harrison C, Thacker C, Catchpole B, Syndercombe Court D (2008) Mitochondrial DNA analysis of domestic dogs in the UK.

Forensic Science International, Genetics Supplement Series 1: 598–599. doi: 10.1016/j.fsigss.2007.10.187

Himmelberger AL, Spear TF, Satkoski JA, George DA, Garnica WT, Malladi VS, Smith DG, Webb KM, Allard MW, Kanthaswamy S (2008) Forensic utility of the mitochondrial hypervariable region 1 of domestic dogs, in conjunction with breed and geographic information. Journal of Forensic Sciences 53: 81–89. doi: 10.1111/j.1556-4029.2007.00615.x

Holland MM, Parsons TJ (1999) Mitochondrial DNA Sequence Analysis - Validation and Use for Forensic Casework. Forensic Science Review 11: 21–50.

Imes DL, Wictum EJ, Allard MW, Sacks BN (2012) Identification of single nucleotide polymorphisms within the mtDNA genome of the domestic dog to discriminate individuals with common HVI haplotypes. Forensic Science International, Genetics 6: 630–639. doi: 10.1016/j.fsigen.2012.02.004

Irwin J, Egyed B, Saunier J, Szamosi G, O'Callaghan J, Padar Z, Parsons TJ (2007) Hungarian mtDNA population databases from Budapest and the Baranya county Roma. International Journal of Legal Medicine 121: 377–383. doi: 10.1007/s00414-006-0128-4

Irwin JA, Parson W, Coble MD, Just RS (2011) mtGenome reference population databases and the future of forensic mtDNA analysis. Forensic Science International, Genetics 5: 222–225. doi: 10.1016/j.fsigen.2010.02.008

Kim KS, Jeong HW, Park CK, Ha JH (2001) Suitability of AFLP markers for the study of genetic relationships among Korean native dogs. Genes & Genetic Systems 76: 243–250. doi: 10.1266/ggs.76.243

Kim KS, Lee SE, Jeong HW, Ha JH (1998) The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome. Molecular Phylogenetics and Evolution 10: 210–220. doi: 10.1006/mpev.1998.0513

Klütsch CF, Seppälä EH, Fall T, Uhlén M, Hedhammar Å, Lohi H, Savolainen P (2011a) Regional occurrence, high frequency but low diversity of mitochondrial DNA haplogroup d1 suggests a recent dog-wolf hybridization in Scandinavia. Animal Genetics 42: 100–103. doi: 10.1111/j.1365-2052.2010.02069.x

Klütsch CF, Seppälä EH, Uhlén M, Lohi H, Savolainen P (2011b) Segregation of point mutation heteroplasmy in the control region of dog mtDNA studied systematically in deep generation pedigrees. International Journal of Legal Medicine 125: 527–535. doi: 10.1007/s00414-010-0524-7

Koban E, Saraç ÇG, Açan SC, Savolainen P, Togan İ (2009) Genetic relationship between Kangal, Akbash and other dog populations. Discrete Applied Mathematics 157: 2335–2340. doi: 10.1016/j.dam.2008.06.040

Kropatsch R, Streitberger K, Schulte-Middelmann T, Dekomien G, Epplen JT (2011) On ancestors of dog breeds with focus on Weimaraner hunting dogs. Journal of Animal Breeding and Genetics 128: 64–72. doi: 10.1111/j.1439-0388.2010.00874.x

Li Y, Li Q, Zhao X, Xie Z, Xu Y (2011) Complete sequence of the Tibetan Mastiff mitochondrial genome and its phylogenetic relationship with other Canids (Canis, Canidae). Animal 5: 18–25. doi: 10.1017/S1751731110001370

Li Y, Zhang YP (2012) High genetic diversity of Tibetan Mastiffs revealed by mtDNA sequences. Chinese Science Bulletin 57: 1483–1487. doi: 10.1007/s11434-012-4995-4

Linacre A, Gusmão L, Hecht W, Hellmann AP, Mayr WR, Parson W, Prinz M, Schneider PM, Morling N (2011) ISFG: recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. Forensic Science International, Genetics 5: 501–505. doi: 10.1016/j.fsigen.2010.10.017

Linacre A, Tobe SS (2011) An overview to the investigative approach to species testing in wildlife forensic science. Investigative Genetics 2: 2. doi: 10.1186/2041-2223-2-2

Melton T, Sikora J, Fernandes V, Pereira L (2011) FidoTyping™ and FidoSearch™: Validation of a forensic canine mitochondrial DNA protocol and a new on-line canid mitochondrial hypervariable region database. Mitotyping Technologies, State College, PA, USA, and Institute of Pathology and Molecular Immunology, Porto, Portugal, http://www.mitotyping.com/59859292811581/lib/59859292811581/_files/FidoTyping.ppt

Muñoz-Fuentes V, Darimont CT, Paquet PC, Leonard JA (2010) The genetic legacy of extirpation and re-colonization in Vancouver Island wolves. Conservation Genetics 11: 547–556. doi: 10.1007/s10592-009-9974-1

Nass MM (1969) Mitochondrial DNA. I. Intramitochondrial distribution and structural relations of single- and double-length circular DNA. Journal of Molecular Biology 42: 521–528. doi: 10.1016/0022-2836(69)90240-X

Okumura N, Ishiguro N, Nakano M, Matsui A, Sahara M (1996) Intra- and interbreed genetic variations of mitochondrial DNA major non-coding regions in Japanese native dog breeds (*Canis familiaris*). Animal Genetics 27: 397–405. doi: 10.1111/j.1365-2052.1996.tb00506.x

Okumura N, Ishiguro N, Nakano M, Matsui A, Shigehara N, Nishimoto T, Sahara M (1999) Variations in mitochondrial DNA of dogs isolated from archaeological sites in Japan and neighbouring islands. Anthropological Science 107: 213–228. doi: 10.1537/ase.107.213

Oskarsson MC, Klütsch CF, Boonyaprakob U, Wilton A, Tanabe Y, Savolainen P (2012) Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia for Australian dingoes and Polynesian domestic dogs. Proceedings of the Royal Society B 279: 967–974. doi: 10.1098/rspb.2011.1395

Pang JF, Klütsch C, Zou XJ, Zhang AB, Luo LY, Angleby H, Ardalan A, Ekström C, Sköllermo A, Lundeberg J, Matsumura S, Leitner T, Zhang YP, Savolainen P (2009) mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. Molecular Biology and Evolution 26: 2849–2864. doi: 10.1093/molbev/msp195

Parra D, Méndez S, Cañón J, Dunner S (2008) Genetic differentiation in pointing dog breeds inferred from microsatellites and mitochondrial DNA sequence. Animal Genetics 39: 1–7. doi: 10.1111/j.1365-2052.2007.01658.x

Parson W, Bandelt HJ (2007) Extended guidelines for mtDNA typing of population data in forensic science. Forensic Science International, Genetics 1: 13–19. doi: 10.1016/j.fsigen.2006.11.003

Parson W, Brandstätter A, Alonso A, Brandt N, Brinkmann B, Carracedo A, Corach D, Froment O, Furac I, Grzybowski T, Hedberg K, Keyser-Tracqui C, Kupiec T, Lutz-Bonengel S, Mevag B, Ploski R, Schmitter H, Schneider P, Syndercombe-Court D, Sørensen E, Thew H, Tully G, Scheithauer R (2004) The EDNAP mitochondrial DNA population

database (EMPOP) collaborative exercises: organisation, results and perspectives. Forensic Science International 139: 215–226. doi: 10.1016/j.forsciint.2003.11.008

Parson W, Dür A (2007) EMPOP–a forensic mtDNA database. Forensic Science International, Genetics 1: 88–92. doi: 10.1016/j.fsigen.2007.01.018

Parson W, Roewer L (2010) Publication of population data of linearly inherited DNA markers in the International Journal of Legal Medicine. International Journal of Legal Medicine 124: 505–509. doi: 10.1007/s00414-010-0492-y

Pereira F, Carneiro J, van Asch B (2010) A Guide for Mitochondrial DNA Analysis in Non-Human Forensic Investigations. The Open Forensic Science Journal 3: 33–44.

Pereira L, Cunha C, Amorim A (2004a) Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database. International Journal of Legal Medicine 118: 132–136. doi: 10.1007/s00414-003-0424-1

Pereira L, van Asch B, Amorim A (2004b) Standardisation of nomenclature for dog mtDNA D-loop: a prerequisite for launching a *Canis familiaris* database. Forensic Science International 141: 99–108. doi: 10.1016/j.forsciint.2003.12.014

Pires AE, Ouragh L, Kalboussi M, Matos J, Petrucci-Fonseca F, Bruford MW (2006) Mitochondrial DNA sequence variation in Portuguese native dog breeds: diversity and phylogenetic affinities. Journal of Heredity 97: 318–330. doi: 10.1093/jhered/esl006

Prieto L, Zimmermann B, Goios A, Rodriguez-Monge A, Paneto GG, Alves C, Alonso A, Fridman C, Cardoso S, Lima G, Anjos MJ, Whittle MR, Montesino M, Cicarelli RM, Rocha AM, Albarrán C, de Pancorbo MM, Pinheiro MF, Carvalho M, Sumita DR, Parson W (2011) The GHEP-EMPOP collaboration on mtDNA population data–A new resource for forensic casework. Forensic Science International, Genetics 5: 146–151. doi: 10.1016/j.fsigen.2010.10.013

Randi E, Lucchini V, Christensen MF, Mucci N, Funk SM, Dolf G, Loeschcke V (2000) Mitochondrial DNA Variability in Italian and East European Wolves: Detecting the Consequences of Small Population Size and Hybridization. Conservation Biology 14: 464–473. doi: 10.1046/j.1523-1739.2000.98280.x

Röck A, Irwin J, Dür A, Parsons T, Parson W (2011) SAM: String-based sequence search algorithm for mitochondrial DNA database queries. Forensic Science International, Genetics 5: 126–132. doi: 10.1016/j.fsigen.2010.10.006

Rothuizen J, de Gouw H, Hellebrekers LJ, Lenstra HA (1995) Variable structures of mitochondrial DNA in dogs. Veterinary Quarterly 17: 22–23. doi: 10.1080/01652176.1995.9694575

Ryabinina OM (2006) Mitochondrial DNA Variation in Asian Shepherd Dogs. Russian Journal of Genetics 42: 748–751. doi: 10.1134/S1022795406070088

Salas A, Bandelt HJ, Macaulay V, Richards MB (2007) Phylogeographic investigations: the role of trees in forensic genetics. Forensic Science International 168: 1–13. doi: 10.1016/j.forsciint.2006.05.037

Salas A, Carracedo A, Macaulay V, Richards M, Bandelt HJ (2005) A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics. Biochemical and biophysical research communications 335: 891–899. doi: 10.1016/j.bbrc.2005.07.161

Sato M, Sato K (2013) Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. Biochimica et biophysica acta 1833: 1979–1984. doi: 10.1016/j.bbamcr.2013.03.010

Saunier JL, Irwin JA, Strouss KM, Ragab H, Sturk KA, Parsons TJ (2009) Mitochondrial control region sequences from an Egyptian population sample. Forensic Science International, Genetics 3: e97–103. doi: 10.1016/j.fsigen.2008.09.004

Savolainen P, Leitner T, Wilton AN, Matisoo-Smith E, Lundeberg J (2004) A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. Proceedings of the National Academy of Sciences of the USA 101: 12387–12390. doi: 10.1073/pnas.0401814101

Savolainen P, Lundeberg J (1999) Forensic evidence based on mtDNA from dog and wolf hairs. Journal of Forensic Sciences 44: 77–81.

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

Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T (2002) Genetic evidence for an East Asian origin of domestic dogs. Science 298: 1610–1613. doi: 10.1126/science.1073906

Scharnhorst G, Kanthaswamy S (2011) An assessment of scientific and technical aspects of closed investigations of canine forensics DNA–case series from the University of California, Davis, USA. Croatian Medical Journal 52: 280–292. doi: 10.3325/cmj.2011.52.280

Schneider PM, Seo Y, Rittner C (1999) Forensic mtDNA hair analysis excludes a dog from having caused a traffic accident. International Journal of Legal Medicine 112: 315–316. doi: 10.1007/s004140050257

Sharma DK, Maldonado JE, Jhala YV, Fleischer RC (2004) Ancient wolf lineages in India. Proceedings of the Royal Society of London B (Supplement) 271: S1–S4. doi: 10.1098/rsbl.2003.0071

Sindičić M, Gomerčić T, Galov A, Arbanasić H, Kusak J, Slavica A, Huber Đ (2011) Mitochondrial DNA control region as a tool for species identification and distinction between wolves and dogs from Croatia. Veterinarski Archiv 81: 249–258.

Smalling BB, Satkoski JA, Tom BK, Szeto WY, Erickson BJ-A, Spear TF, Smith DG, Budowle B, Webb KM, Allard M, Kanthaswamy S (2010) Geographic Differences in Mitochondrial DNA (mtDNA) Distribution Among United States (US) Domestic Dog Populations. The Open Forensic Science Journal 3: 22–32.

Suárez NM, Betancor E, Fregel R, Pestano J (2013) Genetic characterization, at the mitochondrial and nuclear DNA levels, of five Canary Island dog breeds. Animal Genetics. doi: 10.1111/age.12024

Sundqvist AK, Björnerfeldt S, Leonard JA, Hailer F, Hedhammar Å, Ellegren H, Vilà C (2006) Unequal contribution of sexes in the origin of dog breeds. Genetics 172: 1121–1128. doi: 10.1534/genetics.105.042358

Takahasi S, Miyahara K, Ishikawa H, Ishiguro N, Suzuki M (2002) Lineage classification of canine inheritable disorders using mitochondrial DNA haplotypes. The Journal of Veterinary Medical Science 64: 255–259. doi: 10.1292/jvms.64.255

Tsuda K, Kikkawa Y, Yonekawa H, Tanabe Y (1997) Extensive interbreeding occurred among multiple matriarchal ancestors during the domestication of dogs: evidence from inter- and intraspecies polymorphisms in the D-loop region of mitochondrial DNA between dogs and wolves. Genes & Genetic Systems 72: 229–238. doi: 10.1266/ggs.72.229

Tully G, Bär W, Brinkmann B, Carracedo A, Gill P, Morling N, Parson W, Schneider P (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. Forensic Science International 124: 83–91. doi: 10.1016/S0379-0738(01)00573-4

Valière N, Fumagalli L, Gielly L, Miquel C, Lequette B, Poulle M-L, Weber J-M, Arlettaz R, Taberlet P (2003) Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. Animal Conservation 6: 83–92. doi: 10.1017/S1367943003003111

van Asch B, Albarran C, Alonso A, Angulo R, Alves C, Betancor E, Catanesi CI, Corach D, Crespillo M, Doutremepuich C, Estonba A, Fernandes AT, Fernandez E, Garcia AM, Garcia MA, Gilardi P, Gonçalves R, Hernández A, Lima G, Nascimento E, de Pancorbo MM, Parra D, Pinheiro Mde F, Prat E, Puente J, Ramírez JL, Rendo F, Rey I, Di Rocco F, Rodríguez A, Sala A, Salla J, Sanchez JJ, Solá D, Silva S, Pestano Brito JJ, Amorim A (2009) Forensic analysis of dog (*Canis lupus familiaris*) mitochondrial DNA sequences: an inter-laboratory study of the GEP-ISFG working group. Forensic Science International, Genetics 4: 49–54. doi: 10.1016/j.fsigen.2009.04.008

van Asch B, Pereira L, Pereira F, Santa-Rita P, Lima M, Amorim A (2005) MtDNA diversity among four Portuguese autochthonous dog breeds: a fine-scale characterisation. BMC Genetics 6: 37. doi: 10.1186/1471-2156-6-37

Vilà C, Maldonado JE, Wayne RK (1999) Phylogenetic relationships, evolution, and genetic diversity of the domestic dog. Journal of Heredity 90: 71–77. doi: 10.1093/jhered/90.1.71

Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK (1997) Multiple and ancient origins of the domestic dog. Science 276: 1687–1689. doi: 10.1126/science.276.5319.1687

Webb K, Allard M (2010) Assessment of minimum sample sizes required to adequately represent diversity reveals inadequacies in datasets of domestic dog mitochondrial DNA. Mitochondrial DNA 21: 19–31. doi: 10.3109/19401730903532044

Webb KM, Allard MW (2009a) Identification of forensically informative SNPs in the domestic dog mitochondrial control region. Journal of Forensic Sciences 54: 289–304. doi: 10.1111/j.1556-4029.2008.00953.x

Webb KM, Allard MW (2009b) Mitochondrial genome DNA analysis of the domestic dog: identifying informative SNPs outside of the control region. Journal of Forensic Sciences 54: 275–288. doi: 10.1111/j.1556-4029.2008.00952.x

Wetton JH, Higgs JE, Spriggs AC, Roney CA, Tsang CS, Foster AP (2003) Mitochondrial profiling of dog hairs. Forensic Science International 133: 235–241. doi: 10.1016/S0379-0738(03)00076-8

Wilson MR, Allard MW, Monson K, Miller KW, Budowle B (2002a) Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region. Forensic Science International 129: 35–42. doi: 10.1016/S0379-0738(02)00206-2

Wilson MR, Allard MW, Monson KL, Miller KWP, Budowle B (2002b) Further Discussion of the Consistent Treatment of Length Variants in the Human Mitochondrial DNA Control Region. Forensic Science Communications 4.

Wilson MR, Stoneking M, Holland MM, DiZinno JA, Budowle B (1993) Guidelines for the Use of Mitochondrial DNA Sequencing in Forensic Science. Crime Laboratory Digest 20: 68–77.

Yao YG, Bravi CM, Bandelt HJ (2004) A call for mtDNA data quality control in forensic science. Forensic Science International 141: 1–6. doi: 10.1016/j.forsciint.2003.12.004

Yao YG, Salas A, Logan I, Bandelt HJ (2009) mtDNA data mining in GenBank needs surveying. The American Journal of Human Genetics 85: 929–933; author reply 933. doi: 10.1016/j.ajhg.2009.10.023

Zimmermann B, Röck A, Huber G, Krämer T, Schneider PM, Parson W (2011) Application of a west Eurasian-specific filter for quasi-median network analysis: Sharpening the blade for mtDNA error detection. Forensic Science International, Genetics 5: 133–137. doi: 10.1016/j.fsigen.2010.10.003