Int J Legal Med (2004) 118: 249–266
DOI 10.1007/s00414-004-0452-5

ORIGINAL ARTICLE

C. Eichmann · B. Berger · W. Parson

A proposed nomenclature for 15 canine-specific polymorphic STR loci for forensic purposes

Received: 11 December 2003 / Accepted: 7 April 2004 / Published online: 28 May 2004

Abstract We performed a population study on 15 polymorphic STR loci (FH2010, FH2079, PEZ2, VWF.X, FH2054, FH2087Ub, FH2611, WILMS-TF, PEZ12, PEZ15, PEZ6, FH2087Ua, ZUBECA4, ZUBECA6, FH2132) on 131 randomly selected dogs. Alleles were identified and grouped according to their estimated fragment length using fixed allelic bins encompassing one base-pair. The allele assignment was confirmed by sequence analysis of homozygote and cloned heterozygote alleles. In order to develop a uniform repeat-based nomenclature, extensive sequence analysis was performed on a selection of alleles from each STR locus. The proposed nomenclature refers to the internationally recognised recommendations for human-specific STR loci in forensic applications. The 15 canine-specific STR loci were grouped into 3 classes (simple STRs, compound STRs and complex/hypervariable STRs) according to their complexity and variability within the repeat structure. Finally, we evaluated the precision of fragment size estimation on a capillary electrophoresis platform and demonstrated reproducibility of fragment length estimation for single base-pair intermediate alleles.

Keywords *Canis familiaris* · Canine short tandem repeats · STR allele nomenclature · Intermediate alleles · Sequence variants

Introduction

Modern dogs (*Canis familiaris*) are direct descendants of wolves (*Canis lupus*) [1]. It has been suggested that there were several independent domestication events and continued genetic exchange between wolves and dogs during coexistence over a wide geographical range [2, 3, 4, 5, 6]. The earliest canine remains date about 10,000–15,000 years ago. Today, more than 400 dog breeds are known to share people's homes, dog populations have reached 25 million in the United States [7], 4.8 million in Germany and 550,000 in Austria (http://www.starkehunde.com/newsroom/Onlinebuch.pdf). As a consequence of this high abundance and the close integration of dogs into human social life, forensically relevant cases involving dogs, such as accidents or dog attacks, are observed on a regular basis. Forensic approaches to fatal dog attacks and a literature review have been published recently, demonstrating the relevance of this issue in forensic casework [7, 8, 9, 10]. If dog hairs are involved, mtDNA analysis is a proper tool for investigating such cases, as described for example in [11, 12]. The DNA profiling of canine stain material has been described previously [10, 11, 12, 13, 14]. In those cases, the evidentiary material was investigated by means of canine-specific STR markers either with the aid of commercially available kits or by a selection of polymorphic canine STRs from population genetic studies [10, 11, 12, 13, 15, 16, 17, 18]. These studies as well as other population genetic investigations on dogs have not been using a uniform repeat-based nomenclature for the STR alleles [10, 11, 12, 13, 14, 15, 16, 18]. The majority of canine short tandem repeat (STR) markers described in the literature are not yet characterised with respect to their sequence structure. The lack of a uniform harmonised nomenclature makes the application of these markers difficult. Neither comparisons between laboratories nor the establishment of frequency databases are possible. Mostly the alleles were reported by the estimated fragment size as determined by electrophoresis of the PCR products. The designation of fragment lengths has the drawback that data which were generated by one laboratory cannot be directly compared with the results of another laboratory, especially when different primers, alternative chemistry and equipment (e.g. electrophoresis instruments) were used. It is a prerequisite for the forensically implementation of canine STR profiling to introduce a nomenclature system for canine-specific STR

C. Eichmann · B. Berger · W. Parson (✉)
Institute of Legal Medicine, University of Innsbruck,
Müllerstrasse 44,
6020 Innsbruck, Austria
e-mail: Walther.Parson@uibk.ac.at
Tel.: +43-512-5073303
Fax: +43-512-5072764

markers that relates to the internationally accepted and generally adopted human STR allele nomenclature [19, 20, 21, 22]. In order to collect a representative number of alleles for frequency estimations and a detailed survey of the allele sequences, we performed a population study on 131 randomly selected dogs, all privately owned domestic dogs of various breeds, for 15 canine-specific STRs. Extensive sequencing of the canine STR alleles was conducted, which provided the basis for the establishment of an STR nomenclature system based on the number of repeats within the variable sequence region.

Material and methods

DNA samples and PCR

Buccal swabs were taken from 131 randomly selected dogs living in the area surrounding Innsbruck, Austria. DNA was extracted from mouth swabs using the Chelex method [23]. The selection of 15 canine STR markers—the duplicated marker FH2087Ua and FH2087Ub herein is referred to as 2 loci—is listed in Table 1. The markers were amplified in 50 µl assays including 1×PCR buffer II, 2 mM $MgCl_2$, 200 µM each dNTP, 2 U AmpliTaq Gold polymerase (Applied Biosystems AB, Foster City, CA), 12.5 µg BSA (Serva, Heidelberg) and 100 nM each primer (Table 1). Amplification was performed on a Gene Amp PCR System 9600 (Perkin Elmer, Norwalk, CT) comprising initial denaturation at 95°C for 11 min followed by 30 cycles at 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min and final incubation at 72°C for 30 min. Each primer pair was tested with respect to the annealing temperature (Tm) ranging from 58°C to 62°C and 60°C was found to give the best results with regard to multiplex designs involving all 15 markers.

Capillary electrophoresis

Aliquots of 2 µl of the amplification products were combined with 20 µl deionised formamide including 0.4 µl internal size standard (Genescan-500 ROX, AB), heat-denatured at 95°C for 3 min, snap-cooled on ice, and subjected to capillary electrophoresis on an ABI Prism 3100 Genetic Analyzer using POP 4, 36 cm capillary arrays and default instrument settings as recommended by the manufacturer. The data were analysed using GeneScan Analysis version 3.7 and Genotyper version 2.5 (both AB).

In order to examine 1 bp separation of long STR fragments (i.e. >400 bp fragments amplified at the ZUBECA4-locus) the alleles 44, 44.1, 44.2 of ZUBECA4

Table 1 Primer sequences and fluorescent label for the 15 canine STR loci investigated

Marker	Primer	Primer sequence (5'–3')	Label	Reference
FH2010	FH2010 F	AAATGGAACAGTTGAGCATGC	Joe	[26]
	FH2010 R	CCCCTTACAGCTTCATTTTCC		
FH2079	FH2079 F	CAGCCGAGCACATGGTTT	Fam	[26]
	FH2079 R	ATTGATTCTGATATGCCCAGC		
PEZ2	PEZ2 F	TCCTCTCTAACTGCCTATGC	Joe	[31]
	PEZ2 R	GCCCTTGAATATGAACAATGACACTGTATC		
VWF.X	VWF.X F	CTCCCCTTCTCTACCTCCACCTCTAA	Fam	[34]
	VWF.X R	CAGAGGTCAGCAAGGGTACTATTGTG		
FH2054	FH2054 F	GCCTTATTCATTGCAGTTAGGG	Joe	[26]
	FH2054 R	ATGCTGAGTTTTGAACTTTCCC		
FH2087[1]	FH2087 F	GGTCCCCTTTTGCCATAGTGT	Fam	[26]
	FH2087 R	CAACTCCCTCCCTCATTTCA		
FH2611	FH2611 F	GAAGCCTATGAGCCAGATCA	Fam	CRHP[a]
	FH2611 R	TGTTAGATGATGCCTTCCTTCT		
WILMS-TF	WILMS-TF F	CCCAATCTCCAGAGATTTTCC	Joe	[31]
	WILMS-TF R	CCAGTCTCAGCTGTGTCCAA		
PEZ12	PEZ12 F	GTAGATTAGATCTCAGGCAG	Fam	[31]
	PEZ12 R	TAGGTCCTGGTAGGGTGTGG		
PEZ15	PEZ15 F	CAGTACAGAGTCTGCTTATC	Joe	[31]
	PEZ15 R	CTGGGGCTTAACTCCAAGTTC		
PEZ6	PEZ6 F	ATGAGCACTGGGTGTTATAC	Joe	[31]
	PEZ6 R	ACACAATTGCATTGTCAAAC		
ZUBECA6	ZUBECA6 F	GCCATAAGCCCCAAGCCAGCAG	NED	[30]
	ZUBECA6 R	TGCCTCGTCAGCCCCTCTTTCC		
ZUBECA4	ZUBECA4 F	GAGGGCAGGAGTCATAAAATCAT	Joe	[24]
	ZUBECA4 R	GCCCAGGGACAAACAATCTT		
FH2132	FH2132 F	CACTGGGAGTGGAGACTGCT	Joe	[26]
	FH2132 R	TGCACAGCCAACTAGAGGTG		

[1]The locus FH2087 is a duplicated STR [26] and is referred to as two loci in this study.
[a]*CRHP* The Canine Radiation Hybrid Project accessed on http://www-recomgen univ-rennes1 fr/doggy html.

were used. These alleles were amplified in 94 replicates each and were separated as mentioned.

Allele cloning and sequencing

Homozygote alleles were amplified as described. The selected PCR products were purified with ExoSAP-IT (Amersham Biosciences, Uppsala, Sweden) and sequenced using Big Dye Terminator sequencing reagents (version 2.0; AB; primer concentration 250 nM). The cycler protocol involved 25 cycles for 15 s at 94°C, 30 s at 50°C and at 60°C for 4 min following initial incubation at 96°C for 30 s. The sequencing reaction products were purified from residual dye terminators using AutoSeq G-50 columns, (Amersham Biosciences, Uppsala, Sweden). Heterozygote alleles were cloned prior to sequence analysis. Unpurified PCR products were ligated into the pCR 4-TOPO vector and transformed into chemically competent One Shot TOP10 *E. coli* using the TOPO TA cloning kit for sequencing (Invitrogen Life Technologies, Carlsbad, CA) following the manufacturer's recommendations. Transformed cells were incubated at 37°C overnight on agar plates containing 50 µg/ml kanamycin. Small portions of bacterial colonies were picked with sterile needles and subjected directly to Templi Phi DNA amplification (Amersham Biosciences) following the manufacturer's recommendations for sample denaturation (95°C for 3 min) and isothermal multiple displacement rolling circle amplification (30°C for 15 h). Aliquots of 1 µl of heat-inactivated Templi Phi reactions (65°C for 10 min) were used without further manipulation as templates in cycle sequencing reactions (30 cycles comprising 95°C for 15 s, 50°C for 10 s and 60°C for 4 min after an initial denaturation at 95°C for 2 min). Sequencing primers were M13for (5'-GTAAAACGACGGCCAGTGA-3') and M13rev (5'-GGAAACAGCTATGACCATG-3') at a final concentration of 160 nM.

Electrophoresis was carried out on an ABI Prism 3100 Genetic Analyzer using POP6, 50 cm capillary arrays and default instrument settings. Data were analyzed using Sequencing Analysis version 3.7 (AB) and Sequencher version 4.1 (Gene Codes Corporation, Ann Arbor, MI).

Nomenclature

The proposed nomenclature for the 15 canine STR markers investigated herein is based on the number of repeated units and is adopted from the recommendations of the International Society of Forensic Genetics (ISFG) for the nomenclature of human STRs [19, 20, 21, 22]. The nomenclature for markers with a complex hypervariable polymorphic region required an alternative approach: an appropriate starting and end-point of the repeat region was determined and the number of nucleotides therein were divided by four (assuming general tetrameric repeat structure) according to the recommendations proposed in [19]. In some cases however, this strategy led to predominantly intermediate alleles. In these cases we

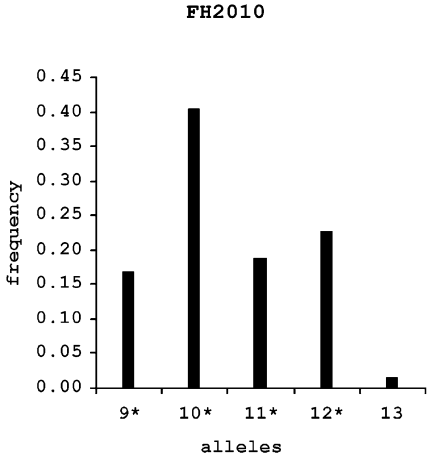
FH2010

FH2010 allele 10 (228 bp)

*AAATGGAACAGTTGAGCATG*CATGTACACCAGAACAT
GAAAAAGC **(ATGA)** $_{10}$AAAAGAGTCGTATTCTTTTAA
ATATTTGTTTTATTGAATTATTTGCACTACTCAATAA
TTTTCTGAGAGTCTGACATTGTTAATAGGAGGAAGCA
TTTCTCTGGAGTTTTAAGGCACAGGCT*GGAAAATGAA*
GCTGTAAGGGG

Sequenced alleles

Allele	(bp)	Repeat structure
9	(224)	45bp **(ATGA)** $_9$ 143bp
10	(228)	45bp **(ATGA)** $_{10}$ 143bp
11	(232)	45bp **(ATGA)** $_{11}$ 143bp
12	(236)	45bp **(ATGA)** $_{12}$ 143bp



allele	n (seq)	bp (min)	bp (max)	d	n	frequency
9	2	223.20	223.39	0.19	44	0.168
10	8	227.23	227.51	0.28	106	0.405
11	2	231.22	231.46	0.24	49	0.187
12	1	235.22	235.50	0.28	59	0.225
13		239.24	239.42	0.18	4	0.015

Fig. 1 FH2010 sequence and electrophoresis precision data as well as allele frequency distribution. *Bold* polymorphic region,*underlined* primer sequence, * sequenced alleles, *bp* base-pairs, *n (seq)* number of sequenced alleles, *bp (min)* and *bp (max)* minimum and maximum observed fragment length of an allele within the study, respectively, *d* bp (max)−bp (min), *n* number of observed alleles

shifted the end-point of the polymorphic region arbitrarily by one or two nucleotides in order to minimise the number intermediate alleles and therefore simplify allele calling.

Results and discussion

Allele assignment

The majority of canine STR markers described in the literature are based on dimeric and tetrameric repeats [24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36]. We selected tetrameric loci only (with the exception of the hexameric locus VWF.X), as dimeric STRs are known to produce stutter artefacts during PCR, which can make allele interpretation difficult [37]. A panel of 15 STR markers was typed in a population sample of 131 randomly selected dogs. The observed fragment lengths of the amplification products were grouped into categories according to their size using allelic bins encompassing 1 bp. Minimal and maximal observed fragment lengths were used in order to assign alleles to categories. Alleles were assigned to the same category when the difference between the minimal and the maximal observed fragment lengths (d-value) was smaller than 0.5 bp. This procedure led to successful allele assignments in the majority of cases (Figs. 1, 2, 3, 4, 5, 6, 7, 8, 9 ,10, 11, 12, 13, 14, 15). Therefore, it was also possible to unambiguously identify intermediate alleles (N.1, N.2 and N.3), which were abundant in complex STR markers (e.g. ZUBECA 4, Fig. 13). These were in part confirmed by sequence analyses. A loss of sizing precision had to be expected for amplicons with fragment sizes exceeding 300 bp [38]. This could potentially result in wrong allele assignments for intermediate alleles differing by a single bp only. In an additional study, we tested the precision of fragment size estimation of high molecular markers, in order to test the sensitivity and reproducibility of single bp resolution with longer fragments. The three sequence-verified ZUBECA 4 alleles 44, 44.1 and 44.2 were analysed in 3 batches consisting of 94 replicate amplifications each. The observed d-values were 0.52, 0.48 and 0.42 (Table 2), demonstrating that single base-pair resolution was generally feasible with fragments exceeding 400 bp in length.

In exceptional cases, the difference between the minimal and maximal apparent fragment lengths of alleles within the same category exceeded 0.5 bp (e.g. FH2132 allele 41, Fig. 15). Sequence analysis of these allele variants revealed divergent repeat structures. We found the following FH2132 allele 41 variants: 5'... $(GAAA)_{12}$ $(GGAA)_{17}$... 3' and 5'... $(GAAA)_{15}(GGAA)_{14}$... 3'. Both variants displayed an identical fragment length consisting of 29 tetrameric repeats at this position. However, their base composition was different, which may have been the cause for the divergent electrophoretic mobilities of the individual sequence variants. This phenomenon has also been described for human-specific markers [38] and underlines the importance of careful interpretation of STR results.

Fig. 2 FH2079 sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

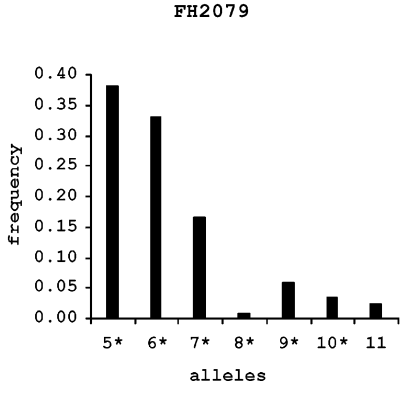
FH2079

FH2079 allele 7 (274 bp)

*CAGCCGAGCACATGGTTT*GTGTTAAACAAATGTTTGAA
TAAATAAATGGATAACTGTTTGGATGGGTGGATGGTGG
ATGGATGGATGGATGATATTCAGTGGATAGATGGATGA
TGGATAGATGGATGATGGATAGATGGATGATTTGGTAG
GT**(GGAT)**$_7$GATGGATGAATAGATGGATGATTCAGTGG
ATGGATGGGTGGATGATGGATGGTTGGATAGATAGATA
ATTA*GCTGGGCATATCAGAATCAAT*

Sequenced alleles

Allele	(bp)	Repeat structure
5	(266)	154bp **(GGAT)**$_5$ 92bp
6	(270)	154bp **(GGAT)**$_6$ 92bp
7	(274)	154bp **(GGAT)**$_7$ 92bp
8	(278)	154bp **(GGAT)**$_8$ 92bp
9	(282)	154bp **(GGAT)**$_9$ 92bp
10	(286)	154bp **(GGAT)**$_{10}$ 92bp



allele	n (seq)	bp (min)	bp (max)	d	n	frequency
5	2	261,81	262,21	0,40	99	0,381
6	4	265,62	266,21	0,59	86	0,331
7	2	269,65	270,09	0,44	43	0,165
8	1	273,96	273,96	-	2	0,008
9	1	277,96	278,26	0,30	15	0,058
10	1	281,96	282,25	0,29	9	0,035
11		285,97	286,05	0,08	6	0,023

Fig. 3 PEZ2 sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

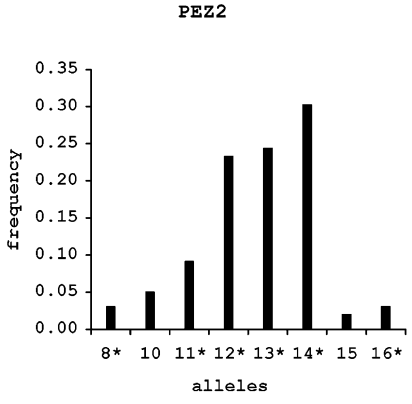
PEZ2

PEZ2 allele 8 (109 bp)

GCCCTTGAATATGAACAATGACACTGTA
TCAA**(GGAA)**$_8$AGAAGGAAGGCAGGCAGA
GGGAGGAGCATAGGCAGTTAGAGAGGA

Sequenced alleles

Allele (bp)		Repeat structure
8	(109)	32bp **(GGAA)**$_8$ 45bp
11	(121)	32bp **(GGAA)**$_{11}$ 45bp
12	(125)	32bp **(GGAA)**$_{12}$ 45bp
13	(129)	32bp **(GGAA)**$_{13}$ 45bp
14	(133)	32bp **(GGAA)**$_{14}$ 45bp
16	(141)	32bp **(GGAA)**$_{16}$ 45bp



allele	n (seq)	bp (min)	bp (max)	d	n	frequency
8	1	105.94	106.04	0.10	8	0.031
10		113.75	114.08	0.33	13	0.050
11	2	117.81	118.02	0.21	24	0.092
12	2	121.78	122.07	0.29	61	0.233
13	1	125.76	126.11	0.35	64	0.244
14	2	129.83	130.20	0.37	79	0.302
15		134.06	134.15	0.09	5	0.019
16	1	138.18	138.27	0.09	8	0.031

Fig. 4 VWF.X sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

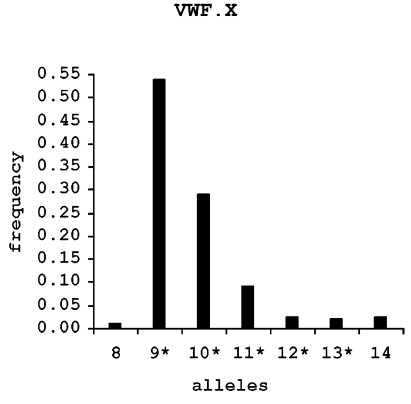
VWF.X

VWF.X allele 9 (157 bp)

CTCCCCTTCTCTACCTCCACCTCTAACTTCA
TCTCATCCCTACCATGTACCTAAAACAGTAG
AAGTCTGG**(AGGAAT)**$_9$GTTGGGTCACAATA
GTACCCTTGCTGACCTCTG

Sequenced alleles

Allele (bp)		Repeat structure
9	(157)	70bp **(AGGAAT)**$_9$ 33bp
10	(163)	70bp **(AGGAAT)**$_{10}$ 33bp
11	(169)	70bp **(AGGAAT)**$_{11}$ 33bp
12	(175)	70bp **(AGGAAT)**$_{12}$ 33bp
13	(181)	70bp **(AGGAAT)**$_{13}$ 33bp



allele	n (seq)	bp (min)	bp (max)	d	n	frequency
8		148.76	148.95	0.19	3	0.012
9	2	155.12	155.45	0.33	137	0.539
10	2	161.39	161.72	0.33	74	0.291
11	1	167.49	167.75	0.26	23	0.091
12	1	173.50	173.70	0.20	6	0.024
13	1	179.61	179.76	0.15	5	0.020
14		185.53	185.73	0.20	6	0.024

Fig. 5 FH2054 sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

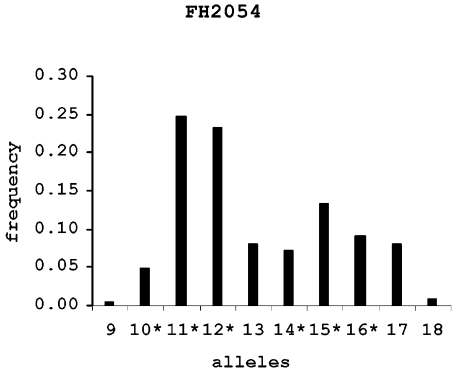
FH2054

FH2054 allele 14 (163 bp)

*ATGCTGAGTTTTGAACTTTCCC*TATCTTCTAG
ATAGATAGATAGAT**(GATA)$_{10}$GTTA(GATA)$_3$**
GATTAGATAATAGCTCCAAAGTTTCTGCTTTT
ATTACAA*CCCTAACTGCAATGAATAAGGC*

Sequenced alleles

Allele	(bp)	Repeat structure
10	(147)	46bp **(GATA)$_{10}$** 61bp
11	(151)	46bp **(GATA)$_{11}$** 61bp
12	(155)	46bp **(GATA)$_8$ GTTA(GATA)$_3$** 61bp
14	(163)	46bp **(GATA)$_{10}$ GTTA(GATA)$_3$** 61bp
15	(167)	46bp **(GATA)$_{11}$ GTTA(GATA)$_3$** 61bp
16	(171)	46bp **(GATA)$_{12}$ GTTA(GATA)$_3$** 61bp



allele	n (seq)	bp (min)	bp (max)	d	n	frequency
9		139.98	139.98	-	1	0.004
10	1	144.36	144.54	0.18	13	0.050
11	4	148.76	149.20	0.44	65	0.248
12	4	153.08	153.54	0.46	61	0.233
13		157.40	157.65	0.25	21	0.080
14	1	161.50	161.75	0.25	19	0.073
15	2	165.66	165.84	0.18	35	0.134
16	1	169.74	169.96	0.22	24	0.092
17		173.74	173.96	0.22	21	0.080
18		177.74	177.93	0.19	2	0.008

Fig. 6 FH2087Ub sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

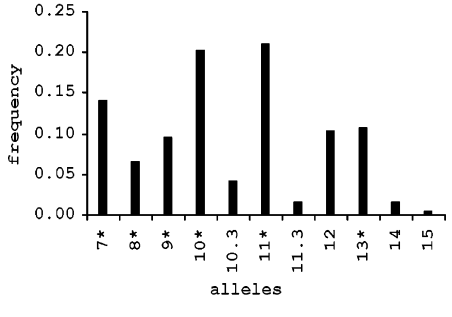
FH2087Ub

FH2087Ub allele 9 (235 bp)

*CTGCCACATTCACTGATGC*ATTTCGCTAAAAATAAG
TAAAATTGTTTTAGCAAAAAAA**(GAAA)$_8$GCAA**GCA
GAGGAAGGAGAGACAGAAAAAGAGACAAAAAAATAA
GGAAAACAAATGCTTGCCTTCTCAGAATAACACATTT
GTCATCATAGACTATACTAACATTAGTATCAGTAAGG
AGAAAGGT*TGAAATGAGGGAGGGAGTTG*

Sequenced alleles

Allele	(bp)	Repeat structure
7	(227)	58bp **(GAAA)$_7$** 141bp
7	(227)	58bp **(GAAA)$_6$ GCAA** 141bp
8	(231)	58bp **(GAAA)$_7$ GCAA** 141bp
9	(235)	58bp **(GAAA)$_8$ GCAA** 141bp
10	(239)	58bp **(GAAA)$_9$ GCAA** 141bp
11	(243)	58bp **(GAAA)$_{10}$ GCAA** 141bp
13	(251)	58bp **(GAAA)$_{13}$** 141bp



allele	n (seq)	bp (min)	bp (max)	d	n	frequency
7	2	223.19	223.43	0.24	37	0.141
8	1	227.08	227.25	0.17	17	0.065
9	1	230.94	231.19	0.25	25	0.095
10	1	234.80	235.08	0.28	53	0.202
10.3		238.08	238.25	0.17	11	0.042
11	2	238.64	238.90	0.26	55	0.210
11.3		241.96	242.15	0.19	4	0.015
12		242.51	242.75	0.24	27	0.103
13	1	246.33	246.56	0.23	28	0.107
14		250.34	250.48	0.14	4	0.015
15		254.30	254.30	-	1	0.004

Fig. 7 FH2611 sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

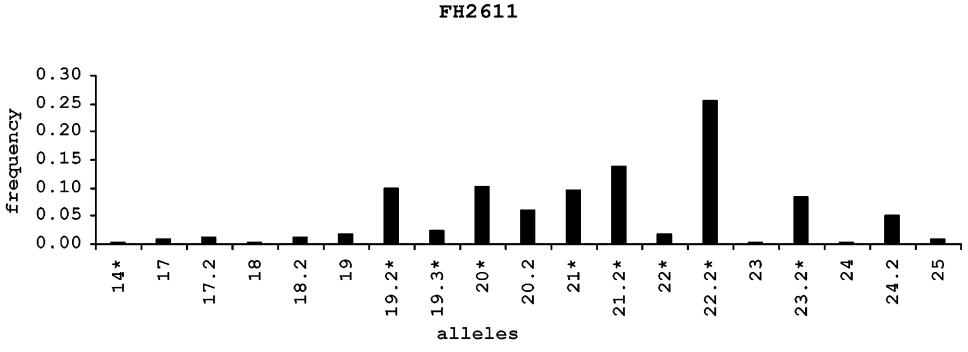
FH2611

FH2611 allele 22 (210 bp)

GAAGCCTATGAGCCAGATCATTCTCAATGTGACCATTATGGCTATTGTTGCTGAGGGTGGAAGT
AGGAA **(GAAA)$_{15}$(GGAA)$_3$(GAGA)$_4$**CGGAGGGTGGAAGGAAGAAGGAAGGAAGGAG*AGAAGGA*
AGGCATCATCTAACA

Sequenced alleles

Allele	(bp)	Repeat structure
14	(178)	69bp **(GAAA)$_7$(GGAA)$_3$(GAGA)$_4$** 53bp
20	(202)	69bp **(GAAA)$_{13}$(GGAA)$_3$(GAGA)$_4$** 53bp
21	(206)	69bp **(GAAA)$_{14}$(GGAA)$_3$(GAGA)$_4$** 53bp
22	(210)	69bp **(GAAA)$_{15}$(GGAA)$_3$(GAGA)$_4$** 53bp
19.2	(200)	69bp **(GAAA)$_{11}$(GGAA)$_5$(GAGA)$_3$(GA)** 53bp
21.2	(208)	69bp **(GAAA)$_{13}$(GGAA)$_5$(GAGA)$_3$(GA)** 53bp
22.2	(212)	69bp **(GAAA)$_{14}$(GGAA)$_5$(GAGA)$_3$(GA)** 53bp
23.2	(216)	69bp **(GAAA)$_{15}$(GGAA)$_5$(GAGA)$_3$(GA)** 53bp
19.3	(201)	69bp **(GAAA)$_6$ A(GAAA)$_6$(GGAA)$_4$(GAGA)$_3$(GA)** 53bp



allele	n (seq)	bp (min)	bp (max)	d	n	frequency	allele	n (seq)	bp (min)	bp (max)	d	n	frequency
14	1	175.63	175.63	-	1	0.004	21	1	202.24	202.42	0.18	25	0.095
17		187.10	187.11	0.01	2	0.008	21.2	2	204.28	204.51	0.23	36	0.137
17.2		188.92	188.94	0.02	3	0.011	22	1	206.13	206.21	0.08	5	0.019
18		190.76	190.76	-	1	0.004	22.2	2	208.12	208.32	0.20	67	0.256
18.2		192.66	192.69	0.03	3	0.011	23		210.11	210.11	-	1	0.004
19		194.67	194.81	0.14	5	0.019	23.2	2	211.92	212.17	0.25	22	0.084
19.2	3	196.34	196.88	0.54	26	0.099	24		214.02	214.02	-	1	0.004
19.3	1	197.54	197.71	0.17	6	0.023	24.2		215.83	216.02	0.19	13	0.050
20	1	198.41	198.60	0.19	27	0.103	25		217.63	217.77	0.14	2	0.008
20.2		200.17	200.59	0.42	16	0.061							

Table 2 Precision data on the electrophoretic separation of variant alleles in a fragment size range of >400 bp. The alleles 44, 44.1 and 44.2 of the marker ZUBECA4 were selected in order to examine reproducible fragment size estimation (94 replicates each)

Allele	Minimal observed fragment length (bp, N=94)	Maximal observed fragment length (bp, N=94)	mean	Difference between minimal and maximal observed fragment lengths (bp)
44	429.66	430.18	**429.98**	0.52
44.1	430.67	431.15	**430.95**	0.48
44.2	431.73	432.15	**431.90**	0.42

Classification of repeat units and allele nomenclature

Sequence analysis of the alleles is a requirement for the establishment of an STR nomenclature based on the number of repeats [19, 20, 21, 22]. We sequenced multiple alleles for each locus in order to attain quantitative data with respect to the true nucleotide length of common alleles and to detect sequence variants (alleles with identical fragment length but different nucleotide composition, e.g. FH2132). Homozygous alleles of each STR

Fig. 8 WILMS-TF sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

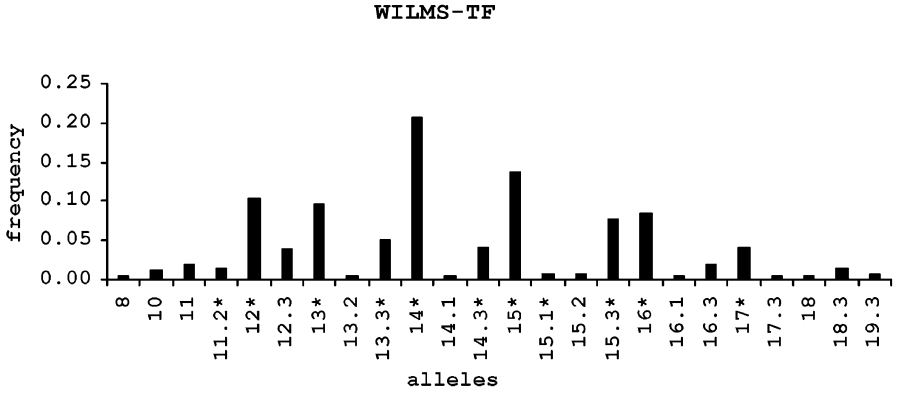
WILMS-TF

WILMS-TF allele 14 (291 bp)

*CCCAATCTCCAGAGATTTTCC*TTTTCTTAAGGGAAAGAAGAAAAAAGAAAGAAAGAAAAGAA
AGAAAA**(GAAA)**$_{14}$AGGAAAGGAAACATCCC*CT***GTTTCTCCTGCAAACCACAGAACAG
TGGGATCCCGCTAGCAATTTGCTGATGATACACTAGTAGTGGTAAAGAACAGGGTGGTGGAG
AACACAATCCGGTTTGTAGGGTGTGATTCACTTTTTCAGTC*TTGGACACAGCTGAGACTGG*

Sequenced alleles

Allele	(bp)	Repeat structure
12	(283)	68bp **(GAAA)**$_{12}$ 167bp
13	(287)	68bp **(GAAA)**$_{13}$ 167bp
14	(291)	68bp **(GAAA)**$_{14}$ 167bp
15	(295)	68bp **(GAAA)**$_{15}$ 167bp
16	(299)	68bp **(GAAA)**$_{16}$ 167bp
17	(303)	68bp **(GAAA)**$_{17}$ 167bp
15.1	(296)	68bp **(GAAA)**$_3$**G(GAAA)**$_{12}$ 167bp
11.2	(281)	68bp **(GAAA)**$_{12}$ 165bp *(DEL)***
13.3	(290)	68bp **(GAAA)**$_{10}$**GAA(GAAA)**$_3$ 167bp
14.3	(294)	68bp **(GAAA)**$_{11}$**GAA(GAAA)**$_3$ 167bp
15.3	(298)	68bp **(GAAA)**$_{12}$**GAA(GAAA)**$_3$ 167bp



allele	n (seq)	bp (min)	bp (max)	d	n	frequency	allele	n (seq)	bp (min)	bp (max)	d	n	frequency
8		263.98	263.98	-	1	0.004	15	5	291.16	291.33	0.17	36	0.137
10		271.79	271.92	0.13	3	0.011	15.1	1	292.34	292.35	0.01	2	0.008
11		275.65	275.88	0.23	5	0.019	15.2		293.14	293.21	0.07	2	0.008
11.2	1	277.52	277.64	0.12	4	0.015	15.3	1	294.05	294.29	0.24	20	0.076
12	2	279.53	279.74	0.21	27	0.103	16	2	295.04	295.27	0.23	22	0.084
12.3		282.47	282.63	0.16	10	0.038	16.1		296.11	296.11	-	1	0.004
13	1	283.41	283.63	0.22	25	0.095	16.3		298.10	298.25	0.15	5	0.019
13.2		285.36	285.36	-	1	0.004	17	1	299.01	299.20	0.19	11	0.042
13.3	1	286.37	286.55	0.18	13	0.050	17.3		302.03	302.03	-	1	0.004
14	5	287.28	287.50	0.22	54	0.206	18		303.02	303.02	-	1	0.004
14.1		288.28	288.28	-	1	0.004	18.3		306.01	306.10	0.09	4	0.015
14.3	1	290.26	290.40	0.14	11	0.042	19.3		309.95	310.02	0.07	2	0.008

**Deletion of two base pairs (CT) in the 3′ flanking region

Fig. 9 PEZ12 sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

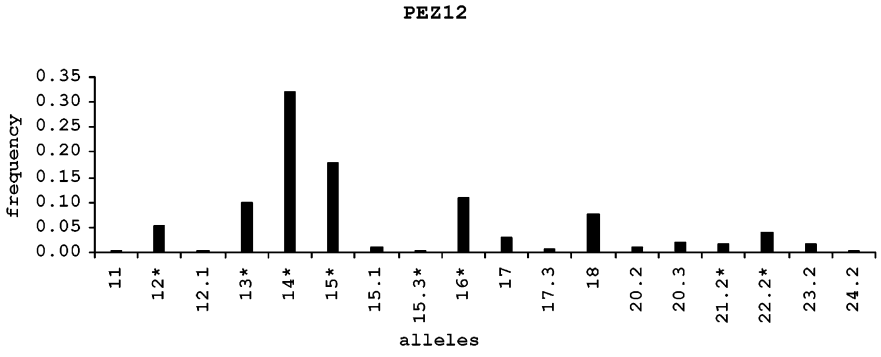
PEZ12

PEZ12 allele 14 (275 bp)

*GTAGATTAGATCTCAGGCAG*TTACAAGCAGTGATTAGAGTTAACTTATACAAAAAAAAAGAAAAAAAG
AAAGAAAGAAAAAGAA **(GAAA)** $_{12}$**GAAGAAAG**ATAAAAAGGATTTGCCAATCAGAAAATAATTTGCTC
AGCAGAAAATAAAGAAAAAGAGAGTCATAGAGGCAAGCATTGCCGAGGTGCACTGCTTAGAGAATGCC
TAGGCCCTGAG*CCACACCCTACCAGGACCTA*

Sequenced alleles

Allele	(bp)	Repeat structure
12	(267)	84bp **(GAAA)** $_{10}$**GAAGAAAG** 135bp
13	(271)	84bp **(GAAA)** $_{11}$**GAAGAAAG** 135bp
14	(275)	84bp **(GAAA)** $_{12}$**GAAGAAAG** 135bp
15	(279)	84bp **(GAAA)** $_{13}$**GAAGAAAG** 135bp
16	(283)	84bp **(GAAA)** $_{14}$**GAAGAAAG** 135bp
21.2	(305)	84bp **(GAAA)** $_{2}$**GAA(GAAA)** $_{11}$**GAA(GAAA)** $_{3}$**GAA(GAAA)** $_{3}$**G** 135bp
22.2	(309)	84bp **(GAAA)** $_{2}$**GAA(GAAA)** $_{12}$**GAA(GAAA)** $_{3}$**GAA(GAAA)** $_{3}$**G** 135bp
15.3	(282)	84bp **(GAAA)** $_{2}$**GAA(GAAA)** $_{10}$**GAA(GAAA)** $_{2}$**G** 135bp



allele	n (seq)	bp (min)	bp (max)	d	n	frequency	allele	n (seq)	bp (min)	bp (max)	d	n	frequency
11		262.78	262.78	-	1	0.004	17		287.15	287.30	0.15	8	0.031
12	1	266.81	267.04	0.23	14	0.054	17.3		290.25	290.46	0.21	2	0.008
12.1		267.93	267.93	-	1	0.004	18		291.18	291.61	0.43	20	0.077
13	1	270.88	271.15	0.27	26	0.100	20.2		301.46	301.66	0.20	3	0.012
14	5	274.90	275.28	0.38	83	0.319	20.3		302.38	302.70	0.32	5	0.019
15	3	278.97	279.31	0.34	46	0.177	21.2	1	305.70	305.87	0.17	4	0.015
15.1		280.03	280.12	0.09	3	0.012	22.2	1	309.72	309.85	0.13	10	0.038
15.3	1	282.15	282.15	-	1	0.004	23.2		313.83	314.03	0.20	4	0.015
16	2	283.09	283.36	0.27	28	0.108	24.2		318.08	318.08	-	1	0.004

Table 3 Precision data on the electrophoretic separation of variant alleles in a fragment size range of >400 bp. The alleles 44, 44.1 and 44.2 of the marker ZUBECA4 were selected in order to examine reproducible fragment size estimation (94 replicates each)

Class	Description	Marker
1	Simple STR loci	FH2010, FH2079, PEZ2, VWF.X
2	Compound STR loci (including intermediate alleles)	FH2054, FH2087Ub, FH2611, WILMS-TF, PEZ12, PEZ15
3	Complex and hypervariable STR loci	PEZ6, FH2087Ua, ZUBECA4, ZUBECA6, FH2132

locus were directly sequenced from the purified PCR product and heterozygote alleles were cloned prior to sequence analysis.

Of the 15 investigated loci, 4 showed a simple repeat structure (FH2010 Fig. 1, FH2079 Fig. 2, PEZ2 Fig. 3 and VWF.X Fig. 4). Allele designation was straightforward in these cases, no single base-pair variants were observed in our population sample. The loci FH2054 (Fig. 5) and FH2087Ub (Fig. 6) consisted of compound repeat units with only few single base-pair length variants (alleles 10.3 and 11.3, FH2087Ub), and were still easy to interpret.

Fig. 10 PEZ15 sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

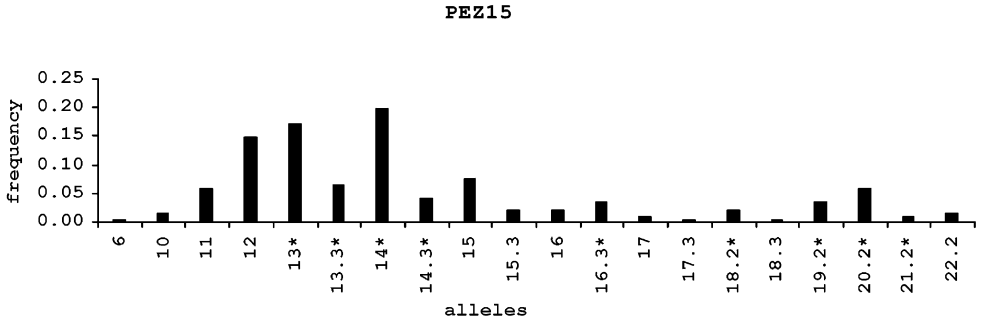
PEZ15

PEZ15 allele 14 (215 bp)

*CAGTACAGAGTCTGCTTATC*CCTCTCCCTCTCCTCTGCCTCTTCCCCCATGTAT
AAATAAAATAAATAAATAAAATCTTAAAATAAAGGAAGAAAGGAAGAAAGAAAGA
(GAAA)$_{13}$GAAGAAAAAGAAAGGAAGAAAGAAAGAACAAAT*GAACTTGGAGTTAAG*
CCCCAG

Sequenced alleles

Allele	(bp)	Repeat structure
13	(211)	109bp **(GAAA)$_{12}$GAAG** 50bp
14	(215)	109bp **(GAAA)$_{13}$GAAG** 50bp
18.2	(233)	109bp **(GAAA)$_4$GAA(GAAA)$_2$GAA(GAAA)$_9$GAAGAAAG** 50bp
19.2	(237)	109bp **(GAAA)$_4$GAA(GAAA)$_2$GAA(GAAA)$_{10}$GAAGAAAG** 50bp
20.2	(241)	109bp **(GAAA)$_4$GAA(GAAA)$_2$GAA(GAAA)$_{11}$GAAGAAAG** 50bp
21.2	(245)	109bp **(GAAA)$_5$GAA(GAAA)$_2$GAA(GAAA)$_{11}$GAAGAAAG** 50bp
13.3	(214)	109bp **(GAAA)$_4$GAA(GAAA)$_8$GAAG** 50bp
14.3	(218)	109bp **(GAAA)$_4$GAA(GAAA)$_9$GAAG** 50bp
16.3	(226)	109bp **(GAAA)$_4$GAA(GAAA)$_{11}$GAAG** 50bp



alleles	n (seq)	bp (min)	bp (max)	d	n	frequency	allele	n (seq)	bp (min)	bp (max)	d	n	frequency
6		178.41	178.41	-	1	0.004	16		216.78	216.89	0.11	5	0.019
10		193.71	193.82	0.11	4	0.015	16.3	1	219.72	219.84	0.12	9	0.034
11		197.52	197.66	0.14	15	0.057	17		220.72	220.74	0.02	2	0.008
12		201.27	201.48	0.21	39	0.149	17.3		223.59	223.59	-	1	0.004
13	3	205.16	205.40	0.24	45	0.172	18.2	2	226.49	226.61	0.12	5	0.019
13.3	1	208.06	208.31	0.25	17	0.065	18.3		227.45	227.45	-	1	0.004
14	2	209.01	209.26	0.25	52	0.198	19.2	1	230.39	230.51	0.12	9	0.034
14.3	1	211.94	212.18	0.24	11	0.042	20.2	1	234.21	234.46	0.25	15	0.057
15		212.86	213.10	0.24	20	0.076	21.2	1	238.15	238.26	0.11	2	0.008
15.3		215.84	216.07	0.23	5	0.019	22.2		242.00	242.14	0.14	4	0.015

The other 9 loci showed a more complex polymorphic region partly including different repeat blocks and incomplete repeat units, which resulted in a relatively high portion of intermediate alleles. Of these, four loci showed an extremely complex repeat structure including dimeric to octomeric motifs as well as insertions of poly-A stretches within the repeat region (FH2087Ua Fig. 12, ZUBECA4 Fig. 13, ZUBECA6 Fig. 14, and FH2132 Fig. 15). In order to set up the above described nomenclature, we traced these complex repeat sequences back to a tetrameric repeat structure. For some loci, the definition of a starting and end-point of the variable repeat region was necessary, in order to assign the complex repeat region to allele categories. In these cases, we defined arbitrary tetrameric STRs by dividing the fragment length of the repeat region by 4, following the recommendations given in [19]. In order to avoid an impractical nomenclature consisting of too many intermediate alleles, we added 1 or 2 bp to the defined polymorphic region, shifting the nomenclature system to mainly integer allele numbers (PEZ12, PEZ15, PEZ6, FH2087Ua, ZUBECA6, and FH2132).

Description of the 15 canine STR loci

In the following section the investigated STR loci are presented in detail. They are described by (1) the general sequence structure including the flanking regions, (2) the structure of the repeat sequences of selected alleles, (3) the

Fig. 11 PEZ6 sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

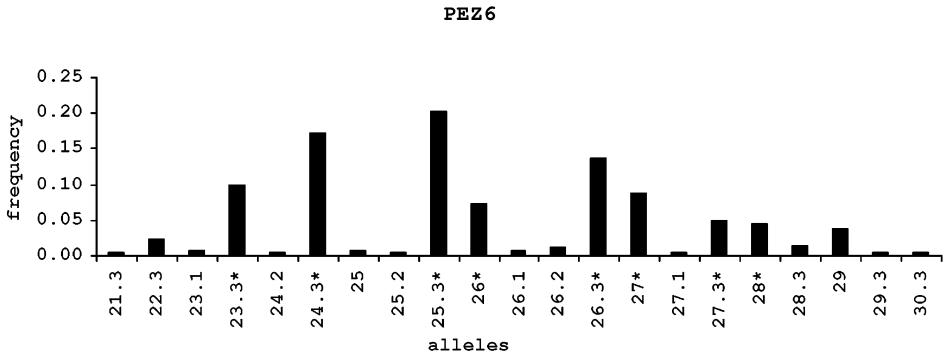
PEZ6

PEZ6 allele 23.3 (176 bp)

*ATGAGCACTGGGTGTTATAC*TATATGTTGGCAAATCGAACTTCAAT
AAAAAAAAGAA (GAAA) $_2$GAAGAAAGAAGGAAAGA (GAAA) $_2$AA (GAAA) $_{12}$A
CCTTTCAAACTTCTA*GTTTGACAATGCAATTGTGT*

Sequenced alleles

Allele	(bp)	Repeat structure
26	(185)	46bp **AAAAAAAGAAGAAAGAAGGAAAGA (GAAA) $_2$AAGAAAGAGAAAGAA (GAAA) $_{14}$ A** 35bp
27	(189)	46bp **AAAAAAAGAAGAAAGAAGGAAAGA (GAAA) $_2$AAGAAAGAGAAAGAA (GAAA) $_{15}$ A** 35bp
28	(193)	46bp **AAAAAAAGAAGAAAGAAGGAAAGA (GAAA) $_2$AAGAAAGAGAAAGAA (GAAA) $_{16}$ A** 35bp
23.3	(176)	46bp **AAAAAAAAGAA (GAAA) $_2$GAAGAAAGAAGGAAAGA (GAAA) $_2$AA (GAAA) $_{12}$ A** 35bp
24.3	(180)	46bp **AAAAAAAAGAA (GAAA) $_2$GAAGAAAGAAGGAAAGA (GAAA) $_2$AA (GAAA) $_{13}$ A** 35bp
25.3	(184)	46bp **AAAAAAAAGAA (GAAA) $_2$GAAGAAAGAAGGAAAGA (GAAA) $_2$AA (GAAA) $_{14}$ A** 35bp
26.3	(188)	46bp **AAAAAAAAGAA (GAAA) $_2$GAAGAAAGAAGGAAAGA (GAAA) $_2$AA (GAAA) $_{15}$ A** 35bp
27.3	(192)	46bp **AAAAAAAAGAA (GAAA) $_2$GAAGAAAGAAGGAAAGA (GAAA) $_2$AA (GAAA) $_{16}$ A** 35bp



alleles	n (seq)	bp (min)	bp (max)	d	n	frequency	alleles	n (seq)	bp (min)	bp (max)	d	n	frequency
21.3		163.9	163.9	0	1	0.004	26.2		182.32	182.37	0.05	3	0.011
22.3		167.75	167.85	0.1	6	0.023	26.3	1	183.18	183.36	0.18	36	0.137
23.1		169.78	169.81	0.03	2	0.008	27	1	184.21	184.4	0.19	23	0.088
23.3	1	171.51	171.72	0.21	26	0.099	27.1		185.19	185.19	0	1	0.004
24.2		174.62	174.62	0	1	0.004	27.3	1	186.98	187.15	0.17	13	0.050
24.3	1	175.45	175.65	0.2	45	0.172	28	1	188.09	188.23	0.14	12	0.046
25		176.6	176.65	0.05	2	0.008	28.3		190.94	190.99	0.05	4	0.015
25.2		178.45	178.45	0	1	0.004	29		191.89	192.06	0.17	10	0.038
25.3	4	179.31	179.53	0.22	53	0.202	29.3		194.72	194.72	0	1	0.004
26	2	180.36	180.66	0.3	19	0.073	30.3		198.54	198.54	0	1	0.004
26.1		181.38	181.4	0.02	2	0.008							

allele requency distribution as observed in our population sample (* indicates that at least one representative of this allele category was sequenced) and (4) precision data of fragment size estimation. If possible and confirmed, we described the STR according to previous published studies. The loci were grouped into three classes according to the complexity of the polymorphic region (Table 3, [39]).

Class 1 STR loci containing a simple repeat motif (FH2010, FH2079, PEZ2, VWF.X)

FH2010 (Fig. 1) The alleles of the locus FH2010 displayed the tetrameric repeat structure ATGA, as described in [26]. All observed alleles clustered into five categories; no intermediate alleles were found. Allele 10 was the most frequent, exceeding a frequency of 0.4. The fragments ranged from 224 bp to 240 bp in length.

FH2079 (Fig. 2) FH2079 displayed the repeat motif GGAT [26]. All of the alleles differed in size by complete

Fig. 12 FH2087Ua sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

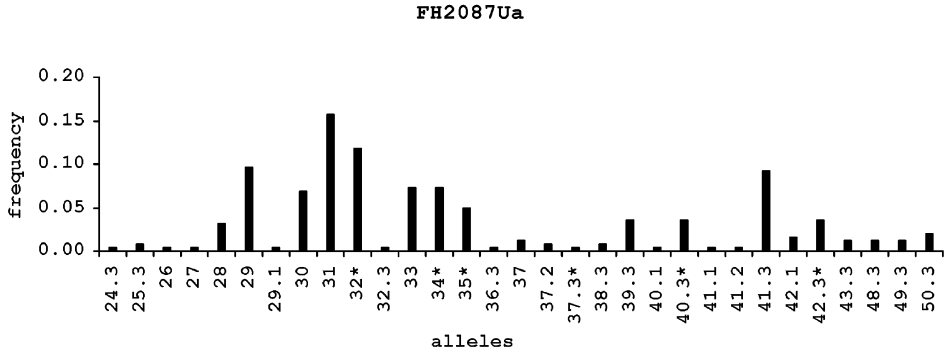
FH2087Ua

FH2087Ua allele 32 (357 bp)

*CTGCCACATTCACTGATGC*ATTTCACAAAAAATAAGAAAATTGTTTTAGGAAAGAAGAAAGAAAGAAA
GAAGAAAGAAAGGAAAGAAGAAAGAAAGAA **(GAAA)** $_{15}$**GAA (GAAA)** $_{2}$**AAAA (GAAA)** $_{11}$**GAA (GA)** $_{2}$**CA**
GAAAGAGAGACAACAAAAAATAAGGAAAACAAATGCTTGCCTTCTCAGAATAACACATTTGTCATC
ATAGACTATACTAACATTAGTATCAGAACCAGTAAGGAAAAAGGC*TGAAATGAGGGAGGGAGTTG*

Sequenced alleles

Allele	(bp)	Repeat structure
32	(357)	98bp **(GAAA)** $_{15}$**GAA (GAAA)** $_{2}$**AAAA (GAAA)** $_{11}$**GAA (GA)** $_{2}$**CA** 131bp
32	(357)	98bp **(GAAA)** $_{14}$**GAA (GAAA)** $_{2}$**AAAA (GA)** $_{2}$ **(GAAA)** $_{11}$**GAA (GA)** $_{2}$**CA** 131bp
34	(365)	98bp **(GAAA)** $_{15}$**GAA (GAAA)** $_{2}$**AAAA (GAAA)** $_{13}$**GAA (GA)** $_{2}$**CA** 131bp
35	(369)	98bp **(GAAA)** $_{16}$**GAA (GAAA)** $_{2}$**AAAA (GAAA)** $_{13}$**GAA (GA)** $_{2}$**CA** 131bp
37.3	(380)	98bp **(GAAA)** $_{13}$**AA (GAAA)** $_{4}$**GAA (GAAA)** $_{3}$**AAA (GAAA)** $_{2}$**GA (GAAA)** $_{11}$**G (GA)** $_{3}$**CA** 131bp
40.3	(392)	98bp **(GAAA)** $_{16}$**AA (GAAA)** $_{4}$**GAAGGAA (GAAA)** $_{2}$**AAA (GAAA)** $_{2}$**GA (GAAA)** $_{11}$**G (GA)** $_{3}$**CA** 131bp
42.3	(400)	98bp **(GAAA)** $_{15}$**AA (GAAA)** $_{4}$**GAA (GAAA)** $_{3}$**AAA (GAAA)** $_{2}$**GA (GAAA)** $_{12}$**GGAAGAAAG (GA)** $_{3}$**CA** 131bp



allele	n (seq)	bp (min)	bp (max)	d	n	frequency	allele	n (seq)	bp (min)	bp (max)	d	n	frequency
24,3		322,71	322,71	-	1	0,004	37,2		372,09	372,09	-	2	0,008
25,3		326,68	326,82	0,14	2	0,008	37,3	1	373,02	373,02	-	1	0,004
26		327,64	327,64	-	1	0,004	38,3		376,81	376,90	0,09	2	0,008
27		331,59	331,59	-	1	0,004	39,3		380,59	380,74	0,15	9	0,034
28		335,39	335,57	0,18	8	0,031	40,1		382,67	382,67	-	1	0,004
29		339,19	339,55	0,36	25	0,095	40,3	1	384,27	384,60	0,33	9	0,034
29,1		340,28	340,28	-	1	0,004	41,1		386,19	386,19	-	1	0,004
30		343,25	343,46	0,21	18	0,069	41,2		387,21	387,21	-	1	0,004
31		347,12	347,43	0,31	41	0,156	41,3		387,99	388,40	0,41	24	0,092
32	2	351,02	351,30	0,28	31	0,118	42,1		389,87	390,19	0,32	4	0,015
32,3		353,90	353,90	-	1	0,004	42,3	1	391,86	392,16	0,30	9	0,034
33		354,87	355,03	0,16	19	0,073	43,3		395,65	395,79	0,14	3	0,011
34	1	358,72	358,87	0,15	19	0,073	48,3		416,01	416,19	0,18	3	0,011
35	1	362,39	362,87	0,48	13	0,050	49,3		419,82	420,17	0,35	3	0,011
36,3		369,32	369,32	-	1	0,004	50,3		423,69	424,09	0,40	5	0,019
37		370,05	370,19	0,14	3	0,011							

tetramer repeat units, no intermediate alleles were found. We observed 86 alleles in category 6, which displayed an unusual high d-value (0.59) compared with other categories. Four homozygote alleles of this category were sequenced, but no sequence variants were observed, which could have explained the different electrophoretic mobility. More sequencing would be necessary to clarify this issue. In our population study we observed 7 alleles ranging between 266 bp and 290 bp in length. Alleles 5 and 6 were most abundant (f>0.3).

PEZ2 (Fig. 3) PEZ2 is described in [31] as a trimeric STR locus. Our population study as well as the sequence analyses of six different alleles clearly demonstrated that the basic repeat motif is tetrameric (GGAA; identical primers as described in [31] were used). We found 8 alleles ranging from 109 bp to 141 bp in our population sample; no intermediate alleles were identified. The most common allele observed was allele 14 with a frequency of 0.3. Allele 9 was not found in this study.

Fig. 13 ZUBECA4 sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

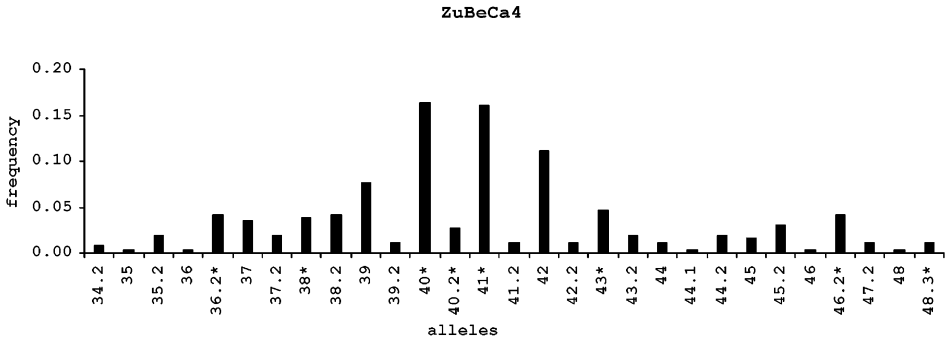
ZUBECA4

ZUBECA4 allele 38 (415 bp)

*GAGGGCAGGAGTCATAAAATCAT*TTTAAAGCGTAGCTATACTATATTGTTCTGTGTGTTGATTTTATGTTCTGCACAAATAAAACTTTAT
TGAAAGAAAGAAAGA **(GAAA)$_6$GAA(GAAA)$_2$GAAGAAAAA(GAAA)$_3$AAAAA(GAAA)$_2$(GAAAA)$_3$AA(GAAA)$_{12}$GA(GAAA)$_4$**AA
GAAGAAAAAGAAAGCTAGCTAGCTAGCTTAGCTGTTTGGGAAAGAGGCCAGTGATGTCAAAGTTAGAATGCACTTATCAAAGTAGACAGT
ATTTAATGTATTGGGAGTAAAATAAGATCAGAGAAAAATCCTTCAC*AAGATTGTTTGTCCCTGGGC*

Sequenced alleles

Allele	(bp)	Repeat structure
38	(415)	105bp **(GAAA)$_6$GAA(GAAA)$_2$GAAGAAAAA(GAAA)$_3$AAAAA(GAAA)$_2$(GAAAA)$_3$AA(GAAA)$_{12}$GA(GAAA)$_4$** 158bp
40	(423)	105bp **(GAAA)$_6$GAA(GAAA)$_2$GAAGAAAAA(GAAA)$_3$AAAAA(GAAA)$_2$(GAAAA)$_3$AA(GAAA)$_{14}$GA(GAAA)$_4$** 158bp
41	(427)	105bp **(GAAA)$_7$GAA(GAAA)$_2$GAAGAAAAA(GAAA)$_4$A(GAAA)$_2$(GAAAA)$_3$AA(GAAA)$_{14}$GA(GAAA)$_4$** 158bp
41	(427)	105bp **(GAAA)$_6$GAA(GAAA)$_2$GAAGAAAAA(GAAA)$_3$AAAAA(GAAA)$_2$(GAAAA)$_3$AA(GAAA)$_{13}$GA(GAAA)$_4$** 158bp
41	(427)	105bp **(GAAA)$_6$GAA(GAAA)$_2$GAAGAAAAA(GAAA)$_3$AAAAA(GAAA)$_2$(GAAAA)$_3$AA(GAAA)$_{15}$GA(GAAA)$_4$** 158bp
43	(435)	105bp **(GAAA)$_6$GAA(GAAA)$_2$GAAGAAAAA(GAAA)$_3$AAAAA(GAAA)$_2$(GAAAA)$_3$AA(GAAA)$_{17}$GA(GAAA)$_4$** 158bp
36.2	(409)	105bp **(GAAA)$_6$GAA(GAAA)$_2$GAAGAAAAA(GAAA)$_3$AAAAA(GAAA)$_2$(GAAAA)$_3$AA(GAAA)$_{15}$** 158bp
40.2	(425)	105bp **(GAAA)$_6$GAAGAAAAA(GAAA)$_3$AAAAA(GAAA)$_4$AGAAAAAA(GAAA)$_{22}$** 158bp
46.2	(449)	105bp **GAA(GAAA)$_8$GAA(GAAA)$_2$GAAGAAAAA(GAAA)$_3$AAAAA(GAAA)$_3$AGAAAAAA(GAAA)$_4$GA(GAAA)$_{19}$** 158bp
48.3	(458)	105bp **GAA(GAAA)$_9$GAA(GAAA)$_2$GAAGAAAAA(GAAA)$_3$AAAAA(GAAA)$_3$AGAAAAAA(GAAA)$_4$AA(GAAA)$_{13}$A(GAAA)$_7$** 158bp



allele	n (Seq)	bp (min)	bp (max)	d	n	frequency	allele	n (Seq)	bp (min)	bp (max)	d	n	frequency
34.2		393.48	393.50	0.02	2	0.008	42		422.09	422.56	0.47	29	0.111
35		395.63	395.63	-	1	0.004	42.2		424.06	424.20	0.14	3	0.011
35.2		397.04	397.37	0.33	5	0.019	43	1	425.96	426.34	0.38	12	0.046
36		399.44	399.44	-	1	0.004	43.2		427.89	428.29	0.40	5	0.019
36.2	1	400.94	401.20	0.26	11	0.042	44		429.97	430.02	0.05	3	0.011
37		402.97	403.16	0.19	9	0.034	44.1		430.95	430.95	-	1	0.004
37.2		404.92	405.01	0.09	5	0.019	44.2		431.98	432.17	0.19	5	0.019
38	1	406.48	407.15	0.67	10	0.038	45		433.92	434.05	0.13	4	0.015
38.2		408.55	408.93	0.38	11	0.042	45.2		435.90	436.03	0.13	8	0.031
39		410.53	410.89	0.36	20	0.076	46		437.85	437.85	-	1	0.004
39.2		412.64	412.70	0.06	3	0.011	46.2	1	439.61	439.97	0.36	11	0.042
40	5	414.36	414.74	0.38	43	0.164	47.2		443.69	443.76	0.07	3	0.011
40.2	1	416.40	416.70	0.30	7	0.027	48		445.55	445.55	-	1	0.004
41	3	418.23	418.62	0.39	42	0.160	48.3	1	448.53	448.53	-	3	0.011
41.2		420.30	420.40	0.10	3	0.011							

VWF.X (Fig. 4) The STR locus VWF.X was found to be based on a hexameric motif (AGGAAT [34]), seven alleles were found in our population study and all alleles differed in size by complete hexameric repeat units only. No intermediate alleles were observed. The fragments ranged from 151 bp to 187 bp in length. Allele 9 was the most abundant with a frequency of 0.5.

Class 2 STR loci containing compound repeat motifs (FH2054, FH2087Ub, FH2611, WILMS-TF, PEZ12, PEZ15)

FH2054 (Fig. 5) The common repeat motif of FH2054 was GATA as also described in [26], 6 out of the 10 alleles found in this population study were sequenced. No intermediate alleles were detected. Alleles 11 and 12 were most common with frequencies of about 0.24. All sequenced alleles with 12 or more repeat units harboured one GTTA-motif, separating the repeat region into two GATA blocks. The fragments ranged from 143 bp to 179 bp in length.

Fig. 14 ZUBECA6 sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

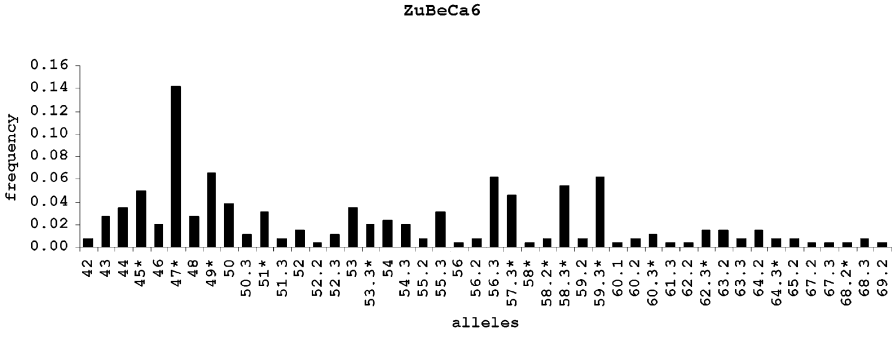
ZUBECA6

ZUBECA6 allele 45 (308 bp)

*GCCATAAGCCCCAAGCCAGCAG*CTATTTGGGGCTGATTAGAACCTGGAATCAGAACAAAGAAAGAGAGAGAGT
(GA)$_{12}$(GAAA)$_{4}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAA(GAAA)$_{2}$GA(GAAA)$_{12}$(GGAA)$_{14}$GAAAGAGAACCTAGAATCGA
TTCATTTTGTTGCATGGG*GGAAAGAGGGGCTGACGAGGCA*

Sequenced alleles

Allele	(bp)	Repeat structure
45	(308)	73bp (GA)$_{12}$(GAAA)$_{4}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAA(GAAA)$_{2}$GA(GAAA)$_{12}$(GGAA)$_{14}$GAAAG 55bp
47	(316)	73bp (GA)$_{12}$(GAAA)$_{4}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAA(GAAA)$_{2}$GA(GAAA)$_{10}$(GGAA)$_{19}$GAAAG 55bp
47	(316)	73bp (GA)$_{12}$(GAAA)$_{4}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAA(GAAA)$_{2}$GA(GAAA)$_{8}$(GGAA)$_{20}$GAAAG 55bp
49	(324)	73bp (GA)$_{12}$(GAAA)$_{4}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAA(GAAA)$_{2}$GA(GAAA)$_{10}$(GGAA)$_{20}$GAAAG 55bp
49	(324)	73bp (GA)$_{12}$(GAAA)$_{4}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAA(GAAA)$_{2}$GA(GAAA)$_{8}$(GGAA)$_{22}$GAAAG 55bp
51	(332)	73bp (GA)$_{12}$(GAAA)$_{4}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAA(GAAA)$_{2}$GA(GAAA)$_{15}$(GGAA)$_{17}$GAAAG 55bp
58	(360)	73bp (GA)$_{12}$(GAAA)$_{4}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAA(GAAA)$_{2}$GA(GAAA)$_{10}$(GGAA)$_{29}$GAAAG 55bp
58.2	(362)	73bp (GA)$_{3}$GAAA(GA)$_{4}$(GAAA)$_{2}$G(GAAA)$_{2}$GAAGA(GAAA)$_{2}$GAA(GAAA)$_{16}$(GGAA)$_{12}$GAAAGA(GAAA)$_{16}$G 55bp
68.2	(402)	73bp (GA)$_{3}$GAAA(GA)$_{4}$(GAAA)$_{2}$G(GAAA)$_{2}$GAAGA(GAAA)$_{2}$GAA(GAAA)$_{19}$(GGAA)$_{14}$GAAAGA(GAAA)$_{21}$G 55bp
53.3	(343)	73bp (GA)$_{3}$GAAA(GA)$_{7}$(GAAA)$_{3}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAAGAAAGA(GAAA)$_{2}$GAA(GAAA)$_{6}$(GGAA)$_{16}$(GAAA)$_{13}$G 55bp
57.3	(359)	73bp (GA)$_{3}$GAAA(GA)$_{4}$(GAAA)$_{3}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAAGA(GAAA)$_{5}$GGAAGAAA(GGAA)$_{18}$GA(GAAA)$_{16}$GAA(GAAA)$_{2}$G 55bp
58.3	(363)	73bp (GA)$_{3}$GAAA(GA)$_{4}$(GAAA)$_{3}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAAGA(GAAA)$_{5}$(GGAA)$_{19}$GA(GAAA)$_{18}$GAA(GAAA)$_{2}$G 55bp
58.3	(363)	73bp (GA)$_{3}$GAAA(GA)$_{4}$(GAAA)$_{3}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAAGA(GAAA)$_{4}$(GGAA)$_{20}$GA(GAAA)$_{18}$GAA(GAAA)$_{2}$G 55bp
59.3	(367)	73bp (GA)$_{3}$GAAA(GA)$_{7}$(GAAA)$_{3}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAAGAAAGA(GAAA)$_{2}$GAA(GAAA)$_{9}$(GGAA)$_{17}$(GAAA)$_{15}$G 55bp
59.3	(367)	73bp (GA)$_{3}$GAAA(GA)$_{4}$(GAAA)$_{3}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAAGA(GAAA)$_{8}$(GGAA)$_{18}$GA(GAAA)$_{20}$GAA(GAAA)$_{2}$G 55bp
60.3	(371)	73bp (GA)$_{3}$GAAA(GA)$_{4}$(GAAA)$_{3}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAAGA(GAAA)$_{8}$(GGAA)$_{22}$GA(GAAA)$_{17}$GAA(GAAA)$_{2}$G 55bp
62.3	(379)	73bp (GA)$_{3}$GAAA(GA)$_{4}$(GAAA)$_{3}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAAGA(GAAA)$_{2}$GAA(GAAA)$_{10}$(GGAA)$_{16}$GAAA(GA)$_{3}$(GAAA)$_{18}$G 55bp
64.3	(387)	73bp (GA)$_{3}$GAAA(GA)$_{7}$(GAAA)$_{3}$A(GAAA)$_{2}$G(GAAA)$_{2}$GAAGAAAGA(GAAA)$_{2}$GAA(GAAA)$_{11}$(GGAA)$_{19}$(GAAA)$_{16}$G55bp



allele	n (Seq)	bp (min)	bp (max)	d	n	frequency	allele	n (Seq)	bp (min)	bp (max)	d	n	frequency
42		296.09	296.22	0.13	2	0.008	56.3		356.31	356.73	0.42	16	0.061
43		300.18	300.44	0.26	7	0.027	57.3	1	360.33	360.63	0.30	12	0.046
44		304.20	304.51	0.31	9	0.034	58	1	360.98	360.98	-	1	0.004
45	1	308.29	308.66	0.37	13	0.050	58.2	1	363.05	363.73	0.68	2	0.008
46		312.37	312.64	0.27	5	0.019	58.3	2	364.36	364.62	0.26	14	0.054
47	2	316.46	316.92	0.46	37	0.142	59.2		367.55	367.72	0.17	2	0.008
48		320.58	320.91	0.33	7	0.027	59.3	2	368.17	368.54	0.37	16	0.061
49	2	324.63	324.94	0.31	17	0.065	60.1		370.68	370.68	-	1	0.004
50		328.65	328.99	0.34	10	0.038	60.2		371.64	371.69	0.05	2	0.008
50.3		332.31	332.50	0.19	3	0.011	60.3	1	372.19	372.47	0.28	3	0.011
51	1	332.83	333.15	0.32	8	0.031	61.3		376.26	376.26	-	1	0.004
51.3		336.09	336.43	0.34	2	0.008	62.2		379.39	379.39	-	1	0.004
52		336.88	337.24	0.36	4	0.015	62.3	1	380.29	380.36	0.07	4	0.015
52.2		339.89	339.89	-	1	0.004	63.2		383.34	383.52	0.18	4	0.015
52.3		340.46	340.63	0.17	3	0.011	63.3		384.18	384.25	0.07	2	0.008
53		340.92	341.37	0.45	9	0.034	64.2		387.26	387.41	0.15	4	0.015
53.3	1	344.34	344.63	0.29	5	0.019	64.3	1	388.06	388.06	-	2	0.008
54		344.99	345.42	0.43	6	0.023	65.2		391.27	391.32	0.05	2	0.008
54.3		348.53	348.64	0.11	5	0.019	67.2		399.18	399.18	-	1	0.004
55.2		351.86	352.02	0.16	2	0.008	67.3		400.00	400.00	-	1	0.004
55.3		352.44	352.76	0.32	8	0.031	68.2	1	403.30	403.30	-	1	0.004
56		353.37	353.37	-	1	0.004	68.3		403.90	403.99	0.09	2	0.008
56.2		355.69	356.00	0.31	2	0.008	69.2		407.21	407.21	-	1	0.004

FH2087Ub (Fig. 6) FH2087U was characterised as a duplicated marker with the two loci FH2087Ua and FH2087Ub [26]. In our study FH2087Ua and FH2087Ub were considered as two different markers, because (1) they belong to different linkage groups (FHCRC Dog Genome Project accessed on http://www.fhcrc.org/science/dog_-

Fig. 15 FH2132 sequence and electrophoresis precision data and allele frequency distribution (comments and abbreviations see Fig. 1)

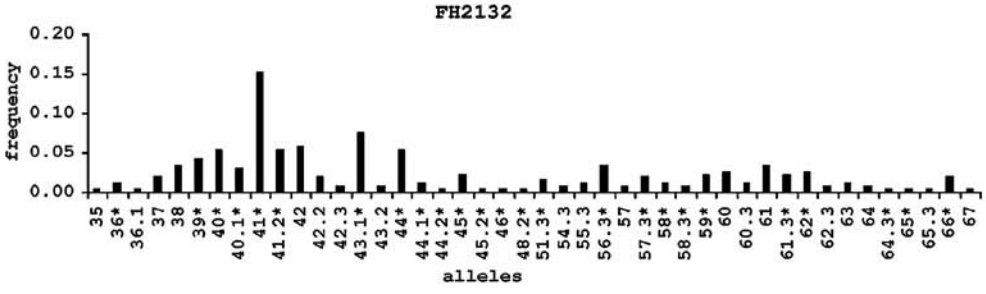
FH2132

FH2132 allele 41 (271 bp)

*CACTGGGAGTGGAGACTGCT*TGGGATTCCCTCTTTTCCTCTCCCTCTACCCCTCCCCTGGCCCTCTCTT
AAAAAGAAAGAA (GAAA)$_{15}$ (GGAA)$_{14}$GAAAGAAAAGAA (GAAA)$_{4}$GAAAAAAGTAAACTTTTTATTACT
TCCACCTCTAGTTGGCTGTGCA

Sequenced alleles

Allele	(bp)	Repeat structure
36	(251)	69bp **AAAAAGAAAGAA (GAAA)$_{13}$ (GGAA)$_{11}$ (GAAA)$_{2}$AGAA (GAAA)$_{5}$AAAG** 38bp
39	(263)	69bp **AAAAAGAAAGAA (GAAA)$_{10}$ (GGAA)$_{17}$ (GAAA)$_{2}$AGAA (GAAA)$_{5}$AAAG** 38bp
40	(267)	69bp **AAAAAGAAAGAA (GAAA)$_{13}$ (GGAA)$_{15}$ (GAAA)$_{2}$AGAA (GAAA)$_{5}$AAAG** 38bp
40	(267)	69bp **AAAAAGAAAGAA (GAAA)$_{15}$ (GGAA)$_{13}$ (GAAA)$_{2}$AGAA (GAAA)$_{5}$AAAG** 38bp
41	(271)	69bp **AAAAAGAAAGAA (GAAA)$_{12}$ (GGAA)$_{17}$ (GAAA)$_{2}$AGAA (GAAA)$_{5}$AAAG** 38bp
41	(271)	69bp **AAAAAGAAAGAA (GAAA)$_{15}$ (GGAA)$_{14}$ (GAAA)$_{2}$AGAA (GAAA)$_{5}$AAAG** 38bp
44	(283)	69bp **AAAAAGAAAGAA (GAAA)$_{15}$ (GGAA)$_{17}$ (GAAA)$_{2}$AGAA (GAAA)$_{5}$AAAG** 38bp
45	(287)	69bp **AAAAAGAAAGAA (GAAA)$_{16}$ (GGAA)$_{17}$ (GAAA)$_{2}$AGAA (GAAA)$_{5}$AAAG** 38bp
46	(291)	69bp **AAAAAGAAAGAA (GAAA)$_{17}$ (GGAA)$_{17}$ (GAAA)$_{2}$AGAA (GAAA)$_{5}$AAAG** 38bp
44.2	(285)	69bp **AAAAAGAAAGAA (GAAA)$_{3}$GA (GAAA)$_{11}$ (GGAA)$_{18}$ (GAAA)$_{2}$AGAA (GAAA)$_{5}$AAAG** 38bp
45.2	(289)	69bp **AAAAAGAAAGAA (GAAA)$_{3}$GA (GAAA)$_{13}$ (GGAA)$_{17}$ (GAAA)$_{2}$AGAA (GAAA)$_{5}$AAAG** 38bp
58	(339)	69bp **AAAAAGAAAGAAGAAAA (GAAA)$_{10}$ (GGAA)$_{15}$ (GAAA)$_{4}$A (GAAA)$_{9}$A (GAAA)$_{14}$AAAAG** 38bp
59	(343)	69bp **AAAAAGAAAGAAGAAAA (GAAA)$_{14}$ (GGAA)$_{13}$ (GAAA)$_{4}$A (GAAA)$_{9}$A (GAAA)$_{13}$AAAAG** 38bp
62	(355)	69bp **AAAAAGAAAGAAGAAAA (GAAA)$_{11}$ (GGAA)$_{15}$ (GAAA)$_{4}$A (GAAA)$_{11}$A (GAAA)$_{15}$AAAAG** 38bp
65	(367)	69bp **AAAAAGAAAGAAGAAAA (GAAA)$_{19}$ (GGAA)$_{12}$ (GAAA)$_{4}$A (GAAA)$_{10}$A (GAAA)$_{14}$AAAAG** 38bp
66	(371)	69bp **AAAAAGAAAGAAGAAAA (GAAA)$_{20}$ (GGAA)$_{12}$ (GAAA)$_{4}$A (GAAA)$_{10}$A (GAAA)$_{14}$AAAAG** 38bp
51.3	(314)	69bp **AAAAAGAAAGAAGAAAA (GAAA)$_{13}$ (GGAA)$_{7}$ (GAAA)$_{3}$AGAA (GAAA)$_{11}$A (GAAA)$_{11}$AAAAG** 38bp
56.3	(334)	69bp **AAAAAGAAAGAAGAAAA (GAAA)$_{12}$ (GGAA)$_{7}$ (GAAA)$_{3}$AGAA (GAAA)$_{14}$A (GAAA)$_{14}$AAAAG** 38bp
57.3	(338)	69bp **AAAAAGAAAGAAGAAAA (GAAA)$_{13}$ (GGAA)$_{7}$ (GAAA)$_{3}$AGAA (GAAA)$_{13}$A (GAAA)$_{15}$AAAAG** 38bp
58.3	(342)	69bp **AAAAAGAAAGAAGAAAA (GAAA)$_{12}$ (GGAA)$_{14}$ (GAAA)$_{4}$AGAA (GAAA)$_{10}$A (GAAA)$_{12}$AAAAG** 38bp
61.3	(354)	69bp **AAAAAGAAAGAAGAAAA (GAAA)$_{12}$ (GGAA)$_{15}$ (GAAA)$_{4}$AGAA (GAAA)$_{11}$A (GAAA)$_{13}$AAAAG** 38bp
64.3	(366)	69bp **AAAAAGAAAGAAGAAAA (GAAA)$_{13}$ (GGAA)$_{16}$ (GAAA)$_{4}$AGAA (GAAA)$_{12}$A (GAAA)$_{13}$AAAAG** 38bp
40.1	(268)	69bp **AAAAAAAA (GAAA)$_{12}$ (GGAA)$_{10}$ (GAAA)$_{2}$AGAA (GAAA)$_{12}$AAAAG** 38bp
43.1	(280)	69bp **AAAAAAAA (GAAA)$_{13}$ (GGAA)$_{10}$ (GAAA)$_{2}$AGAA (GAAA)$_{14}$AAAAG** 38bp
44.1	(284)	69bp **AAAAAAAA (GAAA)$_{13}$ (GGAA)$_{11}$ (GAAA)$_{2}$AGAA (GAAA)$_{14}$AAAAG** 38bp
41.2	(273)	69bp **AAAAAAAA (GAAA)$_{7}$A (GAAA)$_{5}$ (GGAA)$_{12}$ (GAAA)$_{2}$AGAA (GAAA)$_{11}$AAAAG** 38bp
48.2	(301)	69bp **AAAAAAAA (GAAA)$_{7}$A (GAAA)$_{5}$ (GGAA)$_{17}$ (GAAA)$_{2}$AGAA (GAAA)$_{13}$AAAAG** 38bp



allelesn	(Seq)	bp (min)	bp (max)	d	n	frequency	alleles	n	(Seq)	bp (min)	bp (max)	d	n	frequency
35		241.67	241.67	-	1	0.004	51.3	1		306.72	306.74	0.02	4	0.015
36	1	245.59	245.66	0.07	3	0.011	54.3			318.55	318.92	0.37	2	0.008
36.1		246.62	246.62	-	1	0.004	55.3			322.52	323.02	0.50	3	0.011
37		249.40	249.83	0.43	5	0.019	56.3	1		326.34	326.51	0.17	9	0.034
38		253.43	253.79	0.36	9	0.034	57			327.05	327.08	0.03	2	0.008
39	1	257.27	257.74	0.47	11	0.042	57.3	1		330.47	330.77	0.30	5	0.019
40	2	261.22	261.58	0.36	14	0.054	58	1		331.62	331.90	0.28	3	0.011
40.1	1	261.67	261.94	0.27	8	0.031	58.3	1		334.71	334.76	0.05	2	0.008
41	2	265.24	265.79	0.55	40	0.153	59	1		335.58	335.77	0.19	6	0.023
41.2	3	266.80	267.31	0.51	14	0.054	60			339.51	339.74	0.23	7	0.027
42		269.13	269.66	0.53	15	0.057	60.3			342.23	342.75	0.52	3	0.011
42.2		270.74	270.91	0.17	5	0.019	61			343.36	343.74	0.38	9	0.034
42.3		272.00	272.07	0.07	2	0.008	61.3	1		346.65	346.81	0.16	6	0.023
43.1	2	273.11	273.58	0.47	20	0.077	62	1		347.34	347.76	0.42	7	0.027
43.2		274.87	275.25	0.38	2	0.008	62.3			350.45	350.53	0.08	2	0.008
44	1	277.07	277.29	0.22	14	0.054	63			351.38	351.54	0.16	3	0.011
44.1	1	277.45	277.56	0.11	3	0.011	64			355.18	355.24	0.06	2	0.008
44.2	1	279.16	279.16	-	1	0.004	64.3	1		358.09	358.09	-	1	0.004
45	1	281.01	281.36	0.35	6	0.023	65	1		358.87	358.87	-	1	0.004
45.2	1	283.04	283.04	-	1	0.004	65.3			361.85	361.85	-	1	0.004
46	1	284.97	284.97	-	1	0.004	66	1		362.51	362.70	0.19	5	0.019
48.2	1	294.49	294.49	-	1	0.004	67			366.55	366.55	-	1	0.004

genome), (2) their alleles could unambiguously be assigned to one of the two loci, and (3) because of their divergent polymorphic regions. The common repeat motif of FH2087Ub was GAAA, as also described in [26]. Out of 11 alleles found in this population study, 6 were sequenced. The alleles 8, 9, 10 and 11 showed 1 GCAA motif at the 3' end of the repeat block, which we included in the polymorphic region of this marker. Allele 13 consisted of GAAA motifs only. For allele category 7 we found two different repeat variants, 1 sequence showed 7 GAAA motifs, whereas the other consisted of 6 GAAA motifs and 1 GCAA unit. Most of the alleles differed in size by complete tetrameric repeat units. Additionally, we found two N.3 intermediate alleles. The fragments ranged from 227 bp to 259 bp in length.

FH2611 (Fig. 7) The repeat motif of the FH2611 locus is described as tetrameric (FHCRC Dog Genome Project accessed on http://www.fhcrc.org/science/dog_genome and The Canine Radiation Hybrid Project accessed on http://www-recomgen.univ-rennes1.fr/doggy.html). Out of 19 alleles found in this population study, 9 were sequenced and the common repeat structure was found to be $(GAAA)_n$ $(GGAA)_n(GAGA)_n$. Therefore, the allele names were designated by taking all tetramer repeats into consideration. The most common allele found in our population sample was allele 22.2 with a frequency of 0.25. N.2 intermediate alleles showed an additional dinucleotide unit at the 3' end of the repeat block. Allele 19.3 was the only N.3 intermediate allele found in our population sample. This allele showed an insertion of an A within the GAAA block. The fragments ranged from 178 bp to 222 bp in length.

WILMS-TF (Fig. 8) The common repeat motif of the STR locus WILMS-TF was GAAA as described in [31]. Out of 24 observed alleles 11 were sequenced. The intermediate alleles N.1 and N.3 showed incomplete repeat units, whereas the 11.2 intermediate allele comprised a 2 bp deletion in the 3' flanking region. The fragments ranged from 267 bp to 314 bp in length. Allele 14 was the most frequent in our population study (f=0.2).

PEZ12 (Fig. 9) The common repeat motif of the PEZ12 alleles was GAAA as described by [31]. A representative number of intermediate alleles was observed, mainly due to the coexistence of trimeric GAA repeats. In order to achieve uniformity in nomenclature we took advantage of the recommendations in [19], and determined the length of the repeat region, dividing it by four in order to account for the basic tetrameric structure. We set the starting point of the polymorphic region at the first repeat of the variable GAAA block. The end-point of the polymorphic region was arbitrarily shifted 1 bp downstream (now including the G at the 3' end of the repeat region), in order to avoid dominance of intermediate alleles. The fragments ranged from 263 bp to 317 bp in length. Allele 14 was the most abundant with a frequency of 0.32.

PEZ15 (Fig. 10) The common repeat motif of the PEZ15 locus was GAAA, which is also described in [31]. In all N.2 and N.3 intermediate alleles, the GAAA repeat was interrupted by GAA motifs. Similar to the situation for PEZ12, we determined the categories by the length of the polymorphic repeat region and added 1 bp in order to produce predominantly integer alleles. Out of 20 alleles found in this population study 9 were sequenced. The fragments ranged from 183 bp to 249 bp in length. Alleles 13 and 14 were the most abundant (f=0.2) in our population study.

Class 3 STR loci containing complex and hypervariable repeat motifs (PEZ6, FH2087Ua, ZUBECA6, ZUBECA4, FH2132)

Due to the complex nature of the hypervariable region of these STR markers, the nomenclature was generally based on the length of the repeat region and divided by 4, accounting for a tetrameric repeat structure. As described, single basepairs were added to the repeat region in order to favour integer alleles and minimise intermediate alleles.

PEZ6 (Fig. 11) The common repeat motif of the PEZ6 locus was GAAA, which is also described in [31]. Sequence analysis revealed a variable poly-A stretch upstream of the repeat block. The starting point of the polymorphic region was set to the 5' end of the poly-A stretches. The polymorphic region of all sequenced N.3 intermediate alleles differed from the common repeat structure by (1) an additional basepair in the poly-A stretch, (2) the loss of an incomplete repeat unit (GA), and (3) the different consecutive order of the constant repeat motifs with respect to their location within the polymorphic region. (e.g. see alleles 26 and 26.3). The most common allele was 25.3 with a frequency of 0.2. The fragments ranged from 168 bp to 204 bp in length.

FH2087Ua (Fig. 12) The common repeat motif of the locus FH2087Ua was GAAA, as also described in [26]. The polymorphic region contained several GAAA units which were separated by G-rich and A-rich sequences. Out of 31 alleles found in this population study 6 were sequenced. Allele 32 was sequenced twice and showed sequence variability within the second tetranucleotide repeat block. All N.3 intermediate alleles differed from each by a sequence variant within the second tetranucleotide repeat block. The most common allele found in this study was 31 with a frequency of 0.16. The fragments ranged from 328 bp to 432 bp in length.

ZUBECA4 (Fig. 13) The common repeat motif in ZUBECA4 was GAAA, as also described in [24]. Additionally, GA-units, poly-A stretches and pentameric units $(GAAAA)_3$were observed. We set the starting point of the polymorphic region at the first repeat of the first variable GAAA block and defined the end-point with the last GAAA unit. Dolf et al. [24] included a 14 bp fragment

upstream and a 15 bp fragment downstream to define the repeat region in their study. These sequence units were constant in all of our sequenced alleles. Out of 29 alleles found in our population sample 8 were sequenced. We observed sequence variants for alleles with identical fragment length (e.g. allele 41). Allele 41 was sequenced 3 times resulting in 3 variants, which differed within the repeat blocks and the GA-rich sequence units. The most common alleles found in this study were alleles 40 and 41 with a frequency of 0.16. The fragments ranged from 401 bp to 458 bp in length.

ZUBECA6 (Fig. 14) The common repeat motifs of the alleles found in ZUBECA6 were GAAA and GGAA, which was also observed in [30]. The basic repeat structure contained at least one GA-unit as well as several GAAA and one GGAA block. In addition, single nucleotides (A or G) or GA-rich sequence units of variable lengths were inserted. We set the starting point of the polymorphic region at the first repeat of the variable GA block, which is 18 bp downstream of the starting point in [30]. In our data set this 18 bp segment was constant and therefore excluded from the polymorphic region. Out of 46 alleles found in our population sample 14 were sequenced, 4 alleles were sequenced twice and all of them showed sequence variants within the repeat blocks. This emphasises the remarkable variability of this STR locus and indicates that further sequencing is very likely to reveal a great number of new allele variants. The most common allele was 47 with a frequency of 0.14. The fragments ranged from 296 bp to 406 bp in length.

FH2132 (Fig. 15) The common repeat motif of the alleles in FH2132 was GAAA, which was also described in [26]. Additionally, the polymorphic region contained a variable GGAA block, incomplete repeat units and single nucleotides (A) as well as variable GA units and poly-A stretches. Out of 44 alleles 25 were sequenced. All N.1 and N.3 intermediate alleles showed congruent assemblies of their repeat blocks, whereas the integer alleles as well as the N.2 intermediate alleles could be differentiated into two groups. We identified three main clusters of polymorphic regions which distinguished between the groups: (1) Seven integer (36–46) and two N.2 intermediate alleles (44.2 and 45.2) showed the same basic repeat structure with the exception of a GA insertion separating the first variable GAAA block. (2) Five integer alleles (58–66) differed from all N.3 intermediate alleles (51.3–64.3) only by an insertion of a trimeric GAA motif (highlighted in Fig. 15). (3) In contrast to the clusters mentioned, all N.1 and the two N.2 intermediate alleles 41.2 and 48.2 showed an insertion of 3 As in the poly-A stretch.

The high variability of this locus made allele calling difficult in some cases. N.1 intermediate alleles seem to migrate differently compared with others. Alleles 40.1, 43.1 and 44.1 showed mean apparent fragment sizes of 261.8 bp, 273.3 bp and 277.5 bp, respectively. Within these alleles, the differences reflect a tetrameric repeat structure. Allele 42.3 showed a mean apparent fragment size of 272.04 bp, which is only 1 bp (instead of 2) shorter than the apparent size of allele 43.1 (273.3 bp). The reason for this inconsisency may be the sequence differences between the clusters described. This example clearly demonstrates the problems associated with nomenclature issues and single basepair resolution in highly polymorphic STRs. Similar to the ZUBECA loci the various groups and clusters found in the sequenced alleles of FH2132 emphasise the remarkable variability of this STR and indicates that further sequencing will be necessary in order to characterise this locus in more detail.

Conclusions

In this study we characterised simple, compound and complex canine STRs. A selection of these markers may be useful for forensic purposes. For establishing a repeat-based nomenclature, extensive sequence analysis was necessary, especially for the complex markers.

Simple repeat loci and some compound loci may be the markers of choice for routine forensic casework, as they are easy to type and display short fragment lengths, a fact that is important for the analysis of degraded DNA.

The complex markers might be useful in individual cases due to their excellent variability. However, they are difficult to characterise and usually display long fragment lengths, which might limit their use to high quality DNA only. In this study the complex and hypervariable markers were only described to a certain degree because of their high sequence variability. For an advanced characterisation of these markers more sequencing data are needed.

The forensic usefulness of the canine STR markers needs further evaluation in more detail with respect to their discrimination power (e.g. heterozygosity, probability of identity) and their performance for analysing low quantity/degraded DNA, such as encountered in casework (e.g. for the analysis of samples with regard to dog bite injuries). Additionally, the generation of an allelic ladder of the presented loci needs further evaluation so that it can be made available for laboratories interested in the same loci. These studies are underway and will be presented soon.

Acknowledgements The authors would like to thank Mag. vet. med. Volker Büchele, Helena and Paul Scheithauer for bravely collecting the samples. Harald Niederstaetter is kindly acknowledged for valuable discussion.

References

1. Wayne RK (1993) Molecular evolution of the dog family. Trends Genet 9:218–224
2. Flagstad O, Walker CW, Vila C et al. (2003) Two centuries of the Scandinavian wolf population: patterns of genetic variability and migration during an era of dramatic decline. Mol Ecol 12:869–880
3. Vila C, Savolainen P, Maldonado JE et al. (1997) Multiple and ancient origins of the domestic dog. Science 276:1687–1689

4. Vila C, Maldonado JE, Wayne RK (1999) Phylogenetic relationships, evolution, and genetic diversity of the domestic dog. J Hered 90:71–77
5. Vila C, Walker C, Sundqvist AK et al. (2003) Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf-dog hybrids. Heredity 90:17–24
6. Wayne RK, Ostrander EA (1999) Origin, genetic diversity, and genome structure of the domestic dog. Bioessays 21:247–257
7. De Munnynck K, Van de Voorde W (2002) Forensic approach of fatal dog attacks: a case report and literature review. Int J Legal Med 116:295–300
8. Muller S, Flekna G, Muller M, Brem G (1999) Use of canine microsatellite polymorphisms in forensic examinations. J Hered 90:55–56
9. Padar Z, Egyed B, Kontadakis K, Furedi S, Woller J, Zoldag L, Fekete S (2002) Canine STR analyses in forensic practice. Observation of a possible mutation in a dog hair. Int J Legal Med 116:286–288
10. Shutler GG, Gagnon P, Verret G, Kalyn H, Korkosh S, Johnston E, Halverson J (1999) Removal of a PCR inhibitor and resolution of DNA STR types in mixed human-canine stains from a five year old case. J Forensic Sci 44:623–626
11. Savolainen P, Rosen B, Holmberg A, Leitner T, Uhlen M, Lundeberg J (1997) Sequence analysis of domestic dog mitochondrial DNA for forensic use. J Forensic Sci 42:593–600
12. Schneider PM, Seo Y, Rittner C (1999) Forensic mtDNA hair analysis excludes a dog from having caused a traffic accident. Int J Legal Med 112:315–316
13. Ichikawa Y, Takagi K, Tsumagari S et al. (2001) Canine parentage testing based on microsatellite polymorphisms. J Vet Med Sci 63:1209–1213
14. Padar Z, Angyal M, Egyed B, Furedi S, Woller J, Zoldag L, Fekete S (2001) Canine microsatellite polymorphisms as the resolution of an illegal animal death case in a Hungarian zoological gardens. Int J Legal Med 115:79–81
15. Padar Z, Egyed B, Kontadakis K, Zoldag L, Fekete S (2001) Resolution of parentage in dogs by examination of microsatellites after death of putative sire: case report. Acta Vet Hung 49:269–273
16. Zajc I, Sampson J (1996) DNA microsatellites in domesticated dogs: application in paternity disputes. Pflugers Arch 431: R201–202
17. Zajc I, Mellersh C, Kelly EP, Sampson J (1994) A new method of paternity testing for dogs, based on microsatellite sequences. Vet Rec 135:545–547
18. Zajc I, Sampson J (1999) Utility of canine microsatellites in revealing the relationships of pure bred dogs. J Hered 90:104–107
19. Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B (1997) DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics. Int J Legal Med 110:175–176
20. Gill P, Kimpton C, d'Aloja E et al. (1994) Report of the European DNA profiling group (EDNAP)—towards standardisation of short tandem repeat (STR) loci. Forensic Sci Int 65:51–59
21. Gill P, Brinkmann B, d'Aloja E et al. (1997) Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. Forensic Sci Int 87:185–192
22. Gill P, Brenner C, Brinkmann B et al. (2001) DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. Int J Legal Med 114:305–309
23. Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10:506–513
24. Dolf G, Schlapfer J, Switonski M, Stranzinger G, Gaillard C, Schelling C (1998) The highly polymorphic canine microsatellite ZuBeCa4 is localized on canine chromosome 3q15-q18. Anim Genet 29:403–404
25. Dolf G, Schelling C, Stahlberger-Saitbekova N, Fu B, Schlapfer J, Yang F (2000) Seven cosmid-derived canine microsatellites. Anim Genet 31:411–412
26. Francisco LV, Langston AA, Mellersh CS, Neal CL, Ostrander EA (1996) A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. Mamm Genome 7:359–362
27. Holmes NG, Dickens HF, Parker HL, Binns MM, Mellersh CS, Sampson J (1995) Eighteen canine microsatellites. Anim Genet 26:132–133
28. Holmes NG, Dickens HF, Neff MW, Mee JM, Sampson I, Binns MM (1998) Nine canine microsatellites. Anim Genet 29:477
29. Jonasdottir TJ, Dolf G, Sletten M et al. (1999) Five new linkage groups in the canine linkage map. Anim Genet 30:366–370
30. Ladon D, Schelling C, Dolf G, Switonski M, Schlapfer J (1998) The highly polymorphic canine microsatellite ZuBeCa6 is localized on canine chromosome 5q12-q13. Anim Genet 29:466–467
31. Neff MW, Broman KW, Mellersh CS et al. (1999) A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. Genetics 151:803–820
32. Ostrander EA, Sprague GF Jr, Rine J (1993) Identification and characterization of dinucleotide repeat (CA)n markers for genetic mapping in dog. Genomics 16:207–213
33. Richman M, Mellersh CS, Andre C, Galibert F, Ostrander EA (2001) Characterization of a minimal screening set of 172 microsatellite markers for genome-wide screens of the canine genome. J Biochem Biophys Methods 47:137–149
34. Shibuya H, Collins BK, Huang TH, Johnson GS (1994) A polymorphic (AGGAAT)n tandem repeat in an intron of the canine von Willebrand factor gene. Anim Genet 25:122
35. Shibuya H, Collins BK, Collier LL, Huang TH, Nonneman D, Johnson GS (1996) A polymorphic (GAAA)n microsatellite in a canine Wilms tumor 1 (WT1) gene intron. Anim Genet 27:59–60
36. Tiret L, Kessler JL, Bentolila S, Faure S, Bach JM, Weissenbach J, Panthier JJ (2000) Assignation of highly polymorphic markers on a canine purebred pedigree. Mamm Genome 11:703–705
37. Butler J (2001) Forensic DNA typing: biology & technology behind STR markers. Academic Press, San Diego, pp 53–81
38. Sgueglia JB, Geiger S, Davis J (2003) Precision studies using the ABI prism 3100 genetic analyzer for forensic DNA analysis. Anal Bioanal Chem 376:1247–1254
39. Urquhart A, Kimpton CP, Downes TJ, Gill P (1994) Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers. Int J Legal Med 107:13–20