Szent István University
Postgraduate School of Veterinary Science

Investigation of microsatellite polymorphism in degraded canine samples

Theses of Ph.D. dissertation

Written by:

Petra Zenke

2010

Szent István University
Postgraduate School of Veterinary Science

Supervisor:

László Zöldág, DSc
professor
Szent István University, Faculty of Veterinary Science
Institute of Animal Breeding, Nutrition and Laboratory Animal Science
Department of Animal Breeding and Genetics

INTRODUCTION

The investigation of biological samples often requires the analysis of degraded, sometimes even highly degraded materials. As the sample decomposes, the DNA template can become highly fragmented, and the yield of complete target fragments is greatly reduced. The use of polymerase chain reaction (PCR) technology, especially STR systems, has resulted in a significant increase in the succes rate. In such specimens a loss of signal is typically observed with larger-sized STR products (preferential amplification) and the poor amplification of the larger sized loci (300-500 base pairs) in standard multiplex typing kits is commom.

As dogs are widespread and popular pets they are very important from breeding and criminal points of view. In routine dog related caseworks – e.g. dog attacks or traffic accidents – the majority of scene samples are shed hairs and/or degraded remains which often result in a poor or undetectable genetic profile.

The aim of this study was to look for efficient STRs that are robust and informative for successful typing of low amount of highly degraded DNA. The main goals were to:

- reduce the amplicon length of selected known STRs down to 200 bp by designing new primer sequences closely flanking the repetitive region,
- investigate the polymorphism of unknown, potentially efficient STR loci using the short amplicon strategy,
- construct allelic ladders suitable to distinguish the alleles in 1 bp resolutions serving as size standard for each investigated loci,
- create a repeat-based allele nomenclature according to the guidelines for STR markers based on sequencing study,
- establish new multiplexes referred to as „miniplexes” using a combination of redesigned primer pairs,
- carry out population statistical analysis associated with individualization and breeding,
- sensitize the analytical method for a minute amount of DNA (low copy number of DNA),
- test the sufficient application of these miniplexes in both criminal casework and parentage control.

Although in many cases anomalous allelic pattern could be detected during STR-profiling of hair samples, so far only a few investigations have tried to measure the reliability of the detected genotype from single hairs. Due to the intense mitotic activity, somatic mutations are believed to occur with higher frequency in the integument than in other tissues. This, in addition to technical failures adds a further basis for the generation of aberrant DNA-profiles when compared to the actual profile. Body parts that are likely exposed to harmful environmental effects – for example cheek and feet to chemical agents, back to UV-radiation – are expected to produce more frequent mutation events.

The second aim of this study was to measure the success rate of generating a complete genetic profile from single dog hairs in different growth phases collected from different body areas and to find the optimal "low copy number DNA" extraction method. We also planned to define the type and forms of allelic aberrations, to estimate the incidence of these anomalies and to determine the influential factors.

MATERIALS AND METHODS

Primer sequences of PEZ1, PEZ3, PEZ5, PEZ16, PEZ19, PEZ21, FH2054, FH2584, REN124F09, WILMS-TF, vWF.X and SRY were obtained from references and GenBank. In case of loci where amplicon lengths were longer than 200 bp the primers were redesigned. Degree of polymorphism of STR markers were tested on purified DNA samples of unrelated dogs (n = 116) of 75 different breeds with monoplex PCR technique. The amplified fragments were analyzed and sorted by capillary electrophoresis.

Following the monoplex amplification various numbers of alleles from each marker were selected and sequenced from homozygous and heterozygous individuals. For allele calling the reference allelic ladders from sequence verified fragments were constructed mixing the suitable proportion of the monoplex PCR products. Based on the sequence data of the polymorphic region in case of PEZ1, PEZ5, PEZ16, PEZ19, PEZ21, FH2584 and REN124F09 loci a proposed nomenclature for allele calling was developed according to recommendations. In case of PEZ3, FH2054, WILMS-TF and vWF.X loci the nomenclature was previously described.

For allele calling the Genotyper 2.5.2 macros (Applied Biosystems) were modified for each investigated loci and allelic ladders based on the allele sizing.

Based on the number of observed alleles, the heterozygosity (H_{exp}, H_{obs}), the allele frequencies, the values of power of discrimination (PD), power of exclusion (PE) and polymorphism information content (PIC) were calculated. Hardy-Weinberg and linkage equilibrium were tested.

The 12 canine-specific markers - including X- and Y-specific (FH2584 and SRY) markers - were amplified in two miniplex PCR reactions (MP1, MP2). Different fluorescent labels were applied to differentiate the overlapping markers. The sensitivity of two miniplexes was tested in the 2−0.05 ng template range, and I examined practical application of these miniplexes in both criminal casework and parentage control.

Investigation of somatic mutation rate in dog hairs

Plucked dog hairs (n = 600) were collected from four different body sites of a bullmastif donor. After light microscopic examination of the hair roots, the hairs were sorted into three groups based based on different growth and degeneration phases. Buccal swab of the dog individual was taken for reference sample. DNA was isolated from the 0.3-0.5 cm long root end of each hair sample using three different DNA extraction methods: 1. organic extraction [proteinase K digestion, phenol-chlorophorm-isoamylalcohol extraction, Microcon-30 (Millipore) purification and concentration; 2. Qiagen silica-based spin column on BioRobot EZ1; 3. Pellet Paint© Co-Precipitant.

Extracted DNA was quantified and visualized on agarose yield gel with GelRed™ Nucleic Acid Gel Stain (Biotium), and multiplex PCR amplification of ten STR markers (PEZ1, FHC2054, FHC2010, PEZ5, PEZ20, PEZ12, PEZ3, PEZ6, PEZ8, FHC2079) was carried out in 10 µl reaction volume in a GeneAmp PCR System 2720 instrument following the manufacturer’s instructions,. Amplified samples were separated with polyacrylamide gel electrophoresis (PAGE) to estimate capillary electrophoresis loading. Visible PCR products were analyzed on ABI Prism 310 Genetic Analyzer instrument. GeneScan v3.1 and Genotyper v2.5.2. softwares were used for fragment sizing and genotyping.

Samples showing aberrant loci (allele drop-out, imbalance or extra allele) were re-analysed by the same multiplex-, and/or monoplex PCR amplification.

After genotyping, the obtained DNA-profiles were sorted according to the applied DNA purification method, the observed hair phases and the body area from which the hair sample was collected. The pair-wise comparisons of the resulting groups were statistically evaluated using the Yate's correctional χ^2-test of Instat statistical software.

RESULTS

Characterisation of short STR loci

In the cases of five loci where amplicon lengths were longer than 200 bp, the forward and/or reverse primers were redesigned.

Structures and sizes of all allele types of each locus were verified by sequencing. Allele calling was developed based on recommendations. Based on the motif's features the five tetramer (PEZ21, PEZ19, PEZ16, WILMS-TF, FH2584), one pentamer (REN124F09) and one hexamer (vWF.X) loci could be grouped into two categories. The class of each marker depends on the repeat structure and the complexity of the polymorphic region. The PEZ21, PEZ19, vWF.X markers contain simple repetitions while the REN124F09, PEZ16, WILMS-TF, FH2584 loci contain alleles and allelevariants with compound allelestructures.

Representative alleles of loci with previously unregistered sequence data on GenBank (http://blast.ncbi.nlm.nih.gov/) were entered with the following catalog numbers: FJ169896, FJ169897 (PEZ16); FJ169898 (PEZ21); FJ169899 (PEZ19); FJ169900, FJ169901 (REN124F09); FJ169902 (FH2584).

For genotyping the reference allelic ladders from sequence verified fragments were constructed by mixing the suitable proportion of the monoplex PCR products. The allelic ladders were developed from those observed allele ranges with whole repetition number.

The 12 canine-specific markers - including X- and Y-specific (FH2584 and SRY) markers - were amplified in two multiplex PCR reactions (MP1, MP2). The reaction volume of hexaplexes was 20 µl. Primer concentrations were adjusted to balance peak heights within and between dye colors. The applied fluorescent labels (Applied Biosystems) were in MP1: 6-FAM (PEZ1, PEZ16), VIC (PEZ5, REN124F09), NED (PEZ3), PET (PEZ21); and in MP2: 6-FAM (FH2054), VIC (PEZ19, vWF.X), NED (FH2584, SRY), PET (WILMS-TF). The amplifications in both miniplexes were sensitized and tested for a low amount (50 pg) of template DNA in 34 PCR cycles.

Genotypes of 116 dogs were determined on 12 loci and altogether 108 different alleles were detected. In the pooled canine sample the observed allele number on 11 STR loci were 5-24 per locus, and the values of allele frequencies were between 0.4% and 53.9%. Expected heterozygozity (H_{exp}) varied between 0.57-0.92, the values of observed heterozygozity (H_{obs}) varied between 0.31-0.78. However, although some markers (PEZ21, REN124F09) show relatively few alleles, the combined power of discrimination and power of exclusion values are quite high: PD_{comb} = 99,999999999507; PE_{comb} = 99,9953. The polymorphism information content presented by all loci is high enough (PIC = 0.472-0.906).

Except for at three loci (PEZ19, FH2054 and vWF.X) the pooled sample showed significant deviation from Hardy-Weinberg equilibrium at P = 0.0046 significance level (obtained by applying Bonferroni's correction for multiple tests). When testing the linkage between locus pairs only one test (1.82%) showed significant deviation from the state of equilibrium (at P = 0.0046 significance level).

As a result of parentage control using the investigated 11 STR loci the male dog in question was excluded on five markers in one case, six in the other. In the forensic application a whole DNA-profile could be detected from the hair sample (found at a crime scene) with the use of two miniplex reactions. This profile proved to be identical to that from the sample of the suspected dog.

Genetic reproducibility and reliability of DNA-profiles from single dog hairs

Altogether, 524 from the analyzed 600 dog hairs (87.3%) displayed visible PCR product on polyacrylamide gel. In 441 cases (73.5% of total samples) a full genetic profile was observed without locus drop-out and 83 hairs provided partial genotype with more than two locus drop-out. 367 samples resulted in DNA-profiles identical to the reference profile and in 74 cases (12.3% of the total samples) aberrant – imbalanced-, dropped out- and extra allele – DNA-pattern could be detected. The aberrant pattern occurred always at one locus per sample except for in two cases, where a combination of dropped out and extra allele were detected.

Of the three applied DNA-extraction techniques, organic extraction and BioRobot EZ1 methods proved to be the most effective, with which full profiles could be detected in 74.7% and 73.4% (from 486 and 64 hairs, respectively) of the cases, while PelletPaint procedure was less effective with 62.0% from 50 samples.

Only 52.6 % of hair samples (out of 175 hairs) in late catagen phase provided complete genetic profile, while the hairs in anagen or in catagen phases showed 86.7% and 79.4% success rates from 175 and 267 hairs, respectively. There were no significant differences displayed between the hairs originating from different body parts: 74.2% of dorsal hairs (of 260), 67.6% of chop hairs (of 105), 72.9% of front feet hairs (of 107) and 78.0% of hind feet hairs (of 127) provided full DNA-profiles.

The number and the type of aberrant DNA-profiles (which were different from the genotype of the reference sample) were studied using the 441 samples which had no locus drop-out.

Applying BioRobot EZ1, 21.9% of 64 hairs showed anomalous band patterns, which is a significant discrepancy compared to the other two DNA-purification methods (organic:11.1% of 486 hairs, PelletPaint: 12.0% of 50 hairs).

Statistically significant deviations were also detected when comparing the three different hair growth phases: in 14.9% of 175 late catagen hairs, in 13.1% of 267 catagen hairs and in 8.2% of 158 anagen hairs the wrong genotype could be observed. We did not find significant differences in the proportions of the anomalous DNA-profiles with the comparison of the hair samples collected from different body regions.

When re-checking the results by monoplex and/or new multiplex amplification, one sample consistently showed homozygote - instead of heterozygote - genotype at PEZ1 locus.

DISCUSSION

Microsatellite markers for individual identification and for parentage control must have the following attributes: they must be unlinked, highly informative (polymorphic), robust in multiplexes, easy to score and, in addition, should have an internationally accepted allele assignment. The sensitivity of the examination system is essential, especially in the analysis of degraded, low copy DNA samples. The correct interpretation of the results can be attained only with suitable genetic-statistical analyses.

To create an allele nomenclature based on the repeat pattern of polymorphic microsatellites, the determination of sequence data is crucial. Based on sequencing results, I developed an allelic nomenclature which meets the international specifications, and can be used in inter-laboratory practice as well as in forensic cases. My sequencing results, based on the repetition number do not alter the nomenclature proposed by WILMS-TF and vWF.X markers. For the other loci, there was no existing suggested nomenclature.

When selecting markers, the main criterium – in addition to variance – was the size of the microsattelites. Due to this, when needed, I reduced the PCR products length with redesigned oligos differing from earlier published sequences. Doing this, the allele size range does not exceed 200 bp in any of the examination systems (miniplex 1. and miniplex 2.), which results in an increased success rate when performing microsattelite analysis on samples with varying degrees of degradation.

Eleven “mini-STR” loci, and one Y chromosome marker can be amplified in the two, canine-specific PCR multiplexes developed during my thesis work. With the exception of the PEZ3 trimer, the REN124F09 pentamer and the vWF.X hexamer form, all other loci have tetramer repetitions which in contrast to the widely used dinucleotids allow for precise and clear allele calling. The SRY sex-specific marker (located on the Y chromosome) can be deteceted in the case of male dogs in the miniplex 2. reaction, while in the case of females, this sign (peak) cannot be observed.

The correct allele calling for samples of unknown origin can be accomplished by comparing with a selection of reference allelic ladders used as a control. The importance of this is significant in the case of compound STR loci with a large number of allele variants. During assembly of the allelic ladders, I used the most frequently occurring sequenced alleles which contained whole repetions. With the compiled and reproducible allele cocktails and the modified gentyping software the accurate determination of the inter alleles became possible.

The sensitivity of the miniplex 1. and miniplex 2. solutions were tested in 20 μl PCR reaction volume. In both cases 50 pg of DNA template (approximately 10 cells), were sufficient to generate the complete DNA profile.

The minute initial DNA concentration can, however cause stochastic effects in the polymerase chain reaction which can lead to unbalanced loci, allele amplification or even drop-out. For evaluation of the stochastic effects and avoidance of erroneous genotyping, the results should – if possible – be checked with parallel or repeated analyses.

I measured the allele frequency of 11 STR markers for use in individual identification, or parentage control examinations and calculated statistical base values on the basis of 116 canine samples. Those markers also considered effective in forensics generally have high ($H_{(obs)}$ = 0.6–0.8) heterozygosity values, which could be discerened at five loci in the marker set that I used. The significant variance between expected and observed hetrozygosity indicates a lack of genetic balance in the canine group sampled, or in the case of the FH2584 marker, the conspicuous variance could be explained by X chromosome location. The combined power of exclusion (PE) and power of discrimination (PD) calculated for the 11 loci show similarities to the corresponding values applied to human forensics.

The significant deviation of the sampled population from the Hardy-Weinberg and linkage equilibrium can be explained by the heterogenous nature of the group. After all random mating and the associated possibility for genetic drift are limited not only between, but also within breeds (Wahlund effect). The results suggest that we cannot reliably use the genetic database - consisting of data from many different breeds - for statistic analysis without running the risk of misinterpreting the outcome.

In connection with my case study on parentage testing, the practical application of the 12 loci marker set was proved in the determination of questionable paternity in an inbred canine stock, where inbreeding (due to directed crossing), arises more frequently than in the mixed breed group that was used for testing. My forensic case study supports the specifity and sensitivity of the developed miniplexes for successful examination of hairs from crime scenes. In addition to the identification tests, polymorph microsattelite based approaches may be applicable to assign the true genetic status of inbred canine groups, or those showing signs of inbreeding. These results could help in both planning and excuting refreshment of the genetic pool.

Limitation factors of application

Success rate of generating full genotype from dog hair in different growth and degeneration phases correlates with the expected values based on the morphology of the root end – similarly to the human hair examinations. Accordingly the preliminary microscopic checkup of the given hair root type is essential, as generally hair shafts and telogen hairs without adhered hairbulb cells could be successfully genotyped only with the analysis of mitochondrial DNA. As a single hair (particularly shed hairs in telogen phase) may contain an extremely low quantity of nuclear DNA, it is extraordinary important to select the most appropriate DNA-extraction technique. From the three tested DNA-purification procedures that of enzymatic digestion followed by phenol extraction and ultra filtration-concentration (organic method) proved to be the most suitable.

However, there were no significant differences in the results of DNA-analysis of hairs from different parts of the body. During microscopic examination it was found that hairs from the cheek contain a greater proportion of telogen hairs than hairs from other body area. This could complicate the examination of hair samples found within the area of bite traces (in forensic cases).

From 441 apparently complete genetic profiles 74 (16.78%) proved to be divergent − allelic drop-out, imbalance or triallelic pattern − from the original one after the first amplification. To find the reason for these anomalies repeated analysis were carried out. After the new multiplex or monoplex amplification only one sample presented a different genotype − allele 13 on PEZ1 locus − which supports the hypothesis of somatic mutation in the examined dog hair.

NEW SCIENTIFIC RESULTS

1. Polymorphism study of seven microsatellite markers (PEZ21, PEZ16, REN124F09, PEZ19, WILMS-TF, FH2584 and vWF.X) in Hungarian canine stock (116 dogs of 75 different breeds).
2. With new primer sequences closely flanking the repetitive region, PCR-products of five STR markers (PEZ16, REN124F09, PEZ19, WILMS-TF, FH2584) were shortened below 200 basepairs. These primers are specific, stable and usable in multiplex reactions (coamplification).
3. Coamplification of 11 short (90-200 bp), canine specific STRs in two multiplex PCR reactions, completed with the SRY marker (located on the Y chromosome). Miniplex 1.: PEZ1, PEZ5, PEZ3, PEZ21, PEZ16, REN124F09 and Miniplex 2.: PEZ19, WILMS-TF, FH2054, PEZ19, FH2584, SRY.
4. Determination of sequence structure of PEZ21, PEZ16, REN124F09, PEZ19, FH2584 markers and creation of allele nomenclature based on the repeat numbers. Representative alleles of loci were registered into the GenBank with the accession numbers of FJ169896, FJ169897 (PEZ16); FJ169898 (PEZ21); FJ169899 (PEZ19); FJ169900, FJ169901 (REN124F09); FJ169902 (FH2584).
5. Compiling of allelic ladders from sequence verified, reference alleles. Ladders of each locus contain all of the alleles with whole repeat number observed in Hungarian canine stock.
6. For the allele calling the Genotyper 2.5.2 macros were customized for each investigated loci and their allelic ladders to enable semiautomatic interpretation of genotypes.
7. Assessing the efficiency of applied loci in cases of parentage control and individual identification, with the calculation of heterozygosity (H_{exp}, H_{obs}), the values of power of discrimination (PD), power of exclusion (PE) and polymorphism information content (PIC).
8. Sensitize the two miniplex reactions for low copy number of template DNA: approximately 0.05−0.1 ng DNA (from about 10−20 diploid cells) is enough to determine the genetic profile.
9. The sensitivity and the power of discrimination of the two developed multiplexes supported the sufficient application in both criminal casework and parentage control and provides for the possibility of genetic expert opinion supported by DNA tests.

10. Among the collected hairs (n = 600) aberrant genetic patterns were recorded and a somatic mutation event was detected with low frequency (0.2%). The dog studied is a "genetic mosaic" because multiple DNA profiles were present within the body.
11. The significantly high values (16.8%) of anomalous band patterns observed in late catagen hairs during genotyping indicate the need for a careful use of microsatellite markers for low copy number DNA sample analysis.

LIST OF PUBLICATION RELATED TO THE DISSERTATION

Full-text papers published in peer-reviewed journals in English

Zenke P., Egyed B., Zöldág L., Pádár Z.: Population genetic study in Hungarian canine populations using forensically informative STR loci.
Forensic Science International: Genetics (2010) doi:101016/j.fsigen.2010.03.013 [IF: 1.347 expectedly]

Zenke P., Maróti-Agóts Á., Zöldág L., Pádár Zs.: Characterization of the WILMS-TF microsatellite marker in Hungarian dog populations.
Acta Biologica Hungarica (2009) 60(3):329-32 [IF: 0.619]

Zenke P., Egyed B., Szabolcsi Z., Zöldág L., Pádár Zs.: Assessing the frequency of somatic mutation from single dog hairs – Forensic testing of StockMarks Canine I Ver3 kit.
Forensic Science International: Genetics Supplement Series (2008) 1:633-4

Full-text papers published in peer-reviewed journals in Hungarian

Zenke P., Maróti-Agóts Á., Pádár Zs., Gáspárdi A., Komlós I., Zöldág L.: Molecular genetic data to evaluateinbreeding in dog populations.
Magyar Állatorvosok Lapja (2007) 8:484-9 [IF: 0.104]

Zenke P., Pádár Zs., Zöldág L.: Molecular genetics and dog breeding.
Magyar Állatorvosok Lapja (2006) 9:544-50 [IF: 0.155]

Abstract published in peer-reviewed journal in English

Pádár Zs., **Zenke P.**, Egyed B., Ősz K., Kontadakis K., Zöldág L., Fekete S.: STR-Analyse bei Hunden - Forensische Anwendung und Erfahrungen.
Rechtsmedizin (2004) 4:342

Poster or oral presentation on conferences

Zenke P., Pádár Zs., Egyed B., Zöldág L.: STR multiplexes for genotyping low amount of highly degraded canine DNA. *23nd Congress of the International Society for Forensic Genetics, 27-30 May 2008. Ancona,* poster presentation

Zenke P., Leposa T., Pádár Zs., Zöldág L.: Individual identification and parentage control using a hyperpolymorphic microsatellite marker in dog. *I. Gödöllői Állattenyésztési Tudományos Napok, 11-12. April 2008. Gödöllő,* oral presentation

Zenke P., Egyed B., Szabolcsi Z., Zöldág L., Pádár Zs.: Assesing the frequency of somatic mutation from single dog hairs – Forensic testing of StockMarks Canine I Ver3 kit. *22nd Congress of the International Society for Forensic Genetics, 22-25 August, 2007. Copenhagen,* poster presentation

Pádár Zs., **Zenke P**., Egyed B., Ősz K., Kontadakis K., Zöldág L., Fekete S.: STR-Analyse bei hunden forensische anwendung und erfahrungen. *83. Jahrestagung der Deutschen Gesselschaft für Rechtsmedizin, 22-25 September* 2004*. Göttingen,* poster presentation

Pádár Zs., **Zenke P.**, Egyed B., Ősz K., Kontadakis K., Zöldág L., Fekete S.: Experience in the application of commercially available canine-STR kit in Hungarian forensic practice. *82. Jahrestagung der Deutschen Gesselschaft für Rechtsmedizin, 22-25 September 2003. Münster,* poster presentation

Pádár Zs., Egyed B., **Zenke P.**, Ősz K., Zöldág L., Fekete S.: The use of genetic polymorphisms in criminal cynology cases – STR investigations in Hungarien canine population. *V. Magyar Genetikai Kongresszus, 13-15 April 2003. Siófok,* oral presentation

ACKNOWLEDGEMENTS

Special thanks must be extented to my supervisor professor László Zöldág, the head of Department of Animal Breeding and Genetics, for his personal and professional support and his help in the interpretation of the results.

I wish to express my sincere gratitude to Prof. Dr. József Szabó and Dr. István Hullár, the previous and present head of Department of Animal Breeding, Nutrition and Laboratory Animal Science, for providing me facilities to do my research at Faculty of Veterinary Science.

I would like to thank Dr. Zsolt Pádár, the head of Department of Hemogenetics, Institute for Forensic Sciences, who made me notice the importance of forensic animal genetics, inspired me to start on my research, and for his continuous guidance and encouragement.

Dr. Balázs Egyed and his accurate and efficient answers helped me scores of time to solve the problems that occured during the examinations.

I also thank János Woller for developing the allele nomenclature, Zoltán Szabolcsi for cloning and Dr. Viktor Szentgyörgyi for helping me to carry out the statistical analysis.

I wish to thank my previous and present colleagues, Katalin Rajnai, Krisztina Gujdi, dr. Ákos Maróti-Agóts and Judit Üvegesné Nagy for their laboratory and technical assistance.

I would especially thank my reviewers, Dr. Tamás Veresegyházi and Dr. Sándor Füredi, for their time and effort in revising my work and for their constructive criticism and advices to improve the final form of this dissertation.

Other acknowledgements: Grateful thanks to my family and friends for their continuous love and support.