Szent István University
Doctoral (PhD) School of Veterinary Science

Forensic genetic analysis of canine biological remains

PhD defense

Prepared by:
Pádár Zsolt

2006

Szent István University
PhD Doctoral School of Veterinary Science

Thesis advisor and Evaluation Committee

…….
Prof. Dr. Zöldág László CSc
SZIE, Faculty of Veterinary Science, Budapest
Institute of Animal Breeding, Nutrition and Laboratory Animal Science
Division of Animal Breeding and Genetics
Thesis advisor

……
Prof. Dr. Fekete Sándor MTA Doctor
SZIE, Faculty of Veterinary Science, Budapest
Institute of Animal Breeding, Nutrition and Laboratory Animal Science
Division of Laboratory Animal Science
Member of PhD Evaluation Committee

……
Doc. Dr. habil. Veresegyházy Tamás CSc
SZIE, Faculty of Veterinary Science, Budapest
Department of Physiology and Biochemistry
Division of Biochemistry
Member of PhD Evaluation Committee

……..
Pádár Zsolt

1. INTRODUCTION

Forensic genetics based on DNA investigation, formerly considered an interdiscipline, is now, similar to clinical genetics and pharmacogenetics, considered a separate field of genetics. Its challenges, specificity, methods and technical developments as well as its results make it an active contributor to the genetic and legal fields, but cannot be considered to be simply the aggregation of the two fields. Forensic genetics as a distinct field of science uses the methods and scientific information of genetics – e.g. molecular genetics. Its research provides additional, useful information such as molecular structures of polymorphisms, new sequence data, population-genetic values etc. – which in themselves are not the goal of forensic genetics, to other fields of genetics. From the legal point of view, the natural sciences – e.g. genetics – are just tools with the aid of which, criminal hypotheses can be confirmed or rejected. However, professional use of these fields can increase the efficiency of investigation of crimes and legal process.

In addition to identification and individualization of biological material from crime scenes, correlation between crime scenes, personal identification of unknown bodies, and genealogical examination, it can be used for purposes of civil law – parentage – and historical or mass grave analysis. Extending the potential ability of this field to non-human remains could provide a more efficient strategy against crimes and act as an increased criminal deterrent. While use of human microsatellite markers in legal practice is common, in connection with direct or indirect identification of animals, the development of international recommendations has only recently begun.

2. GOAL

The basic goal of my work was, using the model of *Canis familiaris*, to start the process whereby individualization of non human biological materials will become part of forensic genetics.

2.1. Identification of species

Identification of *Canidae* species from samples of unknown origin, using sequence analysis of the homologue section of the cytochrome b gene.

2.2. Analysis of *Canine* specific STR markers in the Hungarian population

Population studies of given loci for paternity analysis, which forensic references have not yet been carried out. These studies are essential for legal applications.

2.2.1. Samples and markers

Samples have been gathered from genetically independent individuals of numerous breeds. The sample types which have forensic relevancy – blood, swabs, hairs etc. – have been analyzed on 10 loci.

2.2.2. Identification of detected fragments, designation and characterization of loci

Accurate definition of the genotype based on a repetitive structure, and survey of possible problems concerning allele size accuracy.

2.2.3. Preparation of allelic ladders, selection of size standards

Accurate genotyping of possibly complex and simple loci using reference alleles and alternative internal size standards.

2.2.4. Semi-automatic analysis

Adaptation of computer program for evaluating detected genetic profiles.

2.3. Population statistical analysis

Carry out population studies, allele frequency, and population comparison studies to provide forensic validation and more exact interpretation of statistic values associated with individualization.

2.3.1. Determination of allele frequencies

2.3.2. Testing of Hardy–Weinberg and linkage equilibrium

2.3.3. Determination of genetic structure and variance

2.4. Testing of different sample types and DNA extraction

Testing of DNA extraction from various biological remains of dogs which putatively have forensic relevance.

2.5. Sensitizing the PCR reaction of cytochrome b and monoplex screening

Development of efficiently sensitive monoplex PCR reactions for successful examination of mixed – human/canine – samples.

2.6. Casework studies

A wider range of possibilities for contribution of expert aid to legal processes as well as efficient use in breeding programs – e.g. parentage control.

3. MATERIALS AND METHODS

3.1. Samples and populations

For population tests, 673 individuals were sampled; consisting of 79 different breeds or breed variance. In the course of the extraction tests, blood, swab, hair, frozen sperm, and frozen biological tissues were tested. The cytochrome b tests were carried out using blood and hair samples, while examination of possible sequence differences was accomplished using DNA extracted from blood of 44 individual animals – 22 German shepherds and 22 Rottweilers. For paternity tests we used blood and hair samples, and forensic tests were conducted using various samples – blood, swab, hair, mixed material – blood/saliva –, and biological tissue remains.

3.2. DNA extraction

DNA recovery from homogeneous samples – liquid blood, saliva, hair etc – was accomplished according to the standard protocols for nuclear and mitochondrial analysis. Extraction from mixed stains was performed using a modified differential lyses (Pádár et al, 1999). Ethidium bromide and agarose electrophoresis were used for the semiquantitative determination of DNA yield.

3.3. Amplification of the fragments to be tested

The 10 autosomal STR were amplified with a StockMarks® Canine I Ver 3 kit and standrad protocol. Amplification of the cytochome sequence has carried out using 3 different primer-pairs and 3 variants of the PCR protocol. Effectiveness of the amplification and separation of fragments was verified with poliacrilamide gel-electrophoresis detected using silver staining.

3.4. Sequencing of the cytochrome b gene segment

Sequencing of amplified and purified fragments was carried out according to standard protocols. I compared the assembled contig to the database of GenBank.

3.5. Genotyping of the STR fragments

The analysis of the STR loci PCR fragments was accomplished using use selected internal standard and validated allele cocktails.

3.5.1. Preparation of allelic ladders

After multiplex tests, products of the selected fragments obtained from monoplex amplification, isolation and cloning must be verified with sequencing. Subsequently, the collected allele mixtures – allele ladders – can be coamplified and reamplified from genotyped DNA samples of individuals. Repetition based allele designation must be developed from the sequenced data.

3.5.2. Comparison of allele-sizing precision

Testing of sizing accuracy was performed using fluorescent size standards with calculated average size and standard deviation values for all allelic ladder fragments.

3.5.3. Modifications to genotyping software

Based on the allele-sizing, the genotyping macros for semiautomatic determination of the alleles should be modified based on the average size values and standard deviation for the internal size standards used.

3.6. Analysis of population statistics

Population statistics – basic genetic values, equilibriums (HWE, LD) – and comparative variance analysis – G, F statistics, AMOVA – was performed on the six population groups.

3.7. Examination of real caseworks

I tested the mixed samples and sensitivity of PCR – with different conditions – from the point of view of casework application. I examined mixed human/canine DNA and real casework samples (Pádár, 2001, 2002, 2003) to simulate the difficulties in analyzing a mixture containing different species.

4. THESES

4.1. Identification of the species

Prior to my work, the given sequence of the cytochrome b gene which contains many variant positions was used for legal identification of samples from unknown species. The primers were constructed for relatively conservative, but human homologue regions. Detection of clear nucleotide position allows comparison of sequences to the homologues in the GenBank database, and allows the taxonomic identification of species. Despite the relative stationarity of primers' annealing site, nucleotide incomplementarity can occur and make the electropherograms and sequence unusable.

4.1.1. Unambiguous sequences

Ambivalent detection can be eliminated in case of other species as well – e.g. Cervidae – and the species' identification can be confirmed using our new reconstructed degenerated primers and PCR conditions.

4.1.2. Detection of sequence variance

Within the significant sequence interval a nucleotide substitution can be observed, the cause of which – single mutation, or possible SNP – requires additional analysis.

4.1.3. Detection of new sequence

Using the previously referred primers with modified PCR conditions, cytochrome b analogue intervals are produced and may be identified as a pseudo-gene located in the nuclear genome, and which may be inserted at several sites in the genome based on their observed heteronucleotide positions. I registered the new sequence on GenBank under catalog number DQ309764.

4.1.4. Possibility for identification of breeds

Since comparison of this new sequence element between breeds – German shepherd/Rottweiler – shows differences, we cannot exclude that these nuclear DNA elements contain breed specific traits, but this must be confirmed with further experiments.

4.2. Individualization – examination of STR polymorphisms

Before the beginning of my study, all of the examined STR markers with the exception of the PEZ20 locus have been described as tetramer repetitions – FHC2010, FHC2054, FHC2079, PEZ3, PEZ5, PEZ6, PEZ8, PEZ12 – or dimer repetitions – PEZ1 – and validated for the

commercial market, but have not been validated for forensic purposes. Clear, structure-based nomenclature, reference allelic ladders for genotyping and allele frequency data were not available, and there was limited technical and methodological knowledge for handling challenging casework.

4.2.1. Description of the repetitive structure of the PEZ20 locus

Sequence analysis revealed the molecular structure of alleles at the given locus (Pádár, 2002).

4.2.2. Clarification of the repetitive structure of the STR loci

The sequence analyses clarified the molecular structures of polymorph markers and discrepancies of repetition structures of PEZ1 and PEZ3 loci to previous references. I registered the unpublished sequence data on GenBank with the following catalog numbers: AF454051, AF454052 (PEZ20), AY375154 (PEZ8), AY375155 (PEZ5), AY375153, AY375156 (PEZ6), AY536266 (PEZ3), AY672136 (PEZ1), AY758357, AY758358 (PEZ12).

4.2.3. Establishment of uniform nomenclature for alleles

Based on the known repetition structures and observed variants I developed a uniform nomenclature for allele calling which is suitable for use in international practice.

4.2.4. Collection of allele ladders

The production of allelic ladders which are semi-automatically genotyped, validated from reference fragments, and tested for sizing accuracy allows for carrying out population studies in the Hungarian canine population.

4.2.5. Allele frequency and population genetics

For the first time in Hungary, I performed population genetic studies based on the observed frequency of STR alleles. Although sampling error in the analyzed population means that the population may not necessarily be representative, we can still declare fixation of significant genetic differences among breed and breed variants. Therefore from a statistical point of view equivalent groups and breeds do not exist in Hungary.

4.2.6. Development of monoplex screening analysis

Sensitization and optimization of monoplex PCR amplification of markers allows for the large scale screening and selection of evidence in casework.

4.2.7. Confirmation of PCR enhancing effect of non-specific DNA

Testing of mixed origin samples confirmed the limitations and ambivalency of cytochrome b analysis as as well as the PCR enhancing effect of nonspecific DNA on nuclear DNA amplification.

4.2.8. Introduction of putative somatic mutation

From a criminalistic perspective, single mutation events cannot be ignored during profiling of hair samples (Pádár, 2002).

4.2.9. Occurrence of identical genetic profiles in different individuals

Occurrence of highly homozygote genetic profile at 10 loci observed during parentage control can indicate a high level of inbreeding in a given population or breeding line. In such cases, distinction between closely related individuals – sire, offspring – using the applied loci may not be possible.

4.2.10. Preparation of forensic genetic expert opinion

As a result of this PhD work, for the first time in Hungary it has become possible to provide an expert forensic genetic opinion supported by DNA based analysis in both parentage control and dog related criminal cases (Pádár, 2001, Pádár, 2002), which have provided relevant information to the legal proceedings abroad.

5. CONCLUSIONS

5.1. Confirmation of species origin

Despite the prior modifications, the primer pairs previously used in forensic practice for quasi-conservative regions of cytochrome b gene have human cytochrome b complementarity, while in canine case, both forward and reverse primers deviate in multiple positions at these sites. The degenerated primers we constructed significantly reduce the deviations of both primers. The annealing insufficiency arising from sequence incomplementarity can be avoided using the new degenerated primer pairs which result in clear, unambiguous sequences. These can be verified and identified in GenBank as canine homologues of the cytochrome b gene located in the mitochondrial genome, which may have possible polymorphism – SNP – in one position.

Although the amplified sequences using the conservative primers and modified PCR conditions show high homologity with *Canidae* taxa, they are not detectable from the pure mitochondrial genome present in hair sheathes.

Due to this, these segments are part of the nuclear genome, which may be inserted as a transpositional element at several sites of the genome, and may multiply and fix as polymorphisms. The probability of this multiplication is supported by the different nucleotides detectable at the same positions, which may have a degree of breed specificity. On the basis of these results, it cannot be excluded that the weaker conservativity and multiplication into the nuclear genome of mitochondrial cytochrome b gene, or gene elements may provide information for exclusion of breed or breed variants as well as possible inclusion. Confirmation of this supposition requires further examination and more consideration.

5.2. Individualization

Polymorph STR markers used in forensics must also satisfy certain special criteria. They must be well defined, identifiable, and their unambiguous detection must be reproducible. They should have a high degree of polymorphism without genetic linkage. From the technical point of view the analytical methods must be sufficiently sensitive and standardizable, and interpretation of the results must be supported by population genetic statistics.

The examined STR loci have 3 levels of structural complexity based on the repetition structure. All of the loci except trimer PEZ3 are based on tetramer repetition. Manifestation of intermediate alleles is based on deletion and addition mutations in flanking region, and partially due to the occurrence of deletion and insertion variants of both flanking and repetitive regions. The largest number of intermediate alleles can be observed at the PEZ6 locus. Sequence variances may appear in both flanking and repetitive regions with same length. On the basis of the sequences, a well-defined allele nomenclature can be developed, which is suitable for data comparison among laboratories and can be recommended for use in international legal applications.

The StockMarks® Canine I Ver 3 kit was basically developed for use in paternity testing. The mixture of 20 primers is not fully optimized for one annealing temperature. The amplification insufficiencies result in allelic and locus dropouts and imbalances, and varying from typical preferential amplification do not correlate to fragment size. The inadequate primer annealing may result in fragments of non-allelic origin. The double peaks caused by imperfect adenilation make the allele calling quite risky in case of interalleles with a single base pair difference, and may lead to false genotyping. In monoplex form using a suitably prepared reaction mix, all loci can be amplified with much less DNA, less number of cycles and without touchdown PCR. During allele sizing, the accuracy relative to the absolute lengths was higher using GS400HD or GS500 standards, but the standard deviation of the relative lengths was significantly less in case of ILS600. Using ILS600 the intermediate alleles can be safely genotyped.

The PEZ6 locus is the most polymorph of the 10 markers, which is caused by its structural complexity and large number of intermediate alleles, although it seems to be possible that appearance of intermediate alleles does not significantly determine the level of polymorphism at given loci. It is, however, more likely that this different level of polymorphism arises from genetic drift and fixation in a given population. The allelic distribution data supports the properties and differences among directly manipulated breeding populations. The deviations of populations from the population equilibriums may explained by complex reasons, the precise determination of which require studies of a larger number of population samples with certified pedigrees. The population structure and variance on a phenotype and molecular level are both characterized by a well defined heterogeneity. Directly manipulated breeding, in contrast to panmixis, can produce an inbred,

highly homozygote population. The differences can be expressed not only among breeds, but also among subpopulations – breeding lines – of the same breed, and consequently, statistical concordance, or replacement of different populations – breeds – is incorrect. Breed uniqueness on a marker level has not been discovered by the recent analyses. The requirement for further population studies in collaboration with breeders and breeding associations is supported by both the legal need for greater effectiveness and the interests of certified breeders.

5.3. Legal applications

The local molecular genetic analyses of animal polymorphism for parentage control is currently usually characterized by the use of custom developed marker sets, analytical and reference systems, identification codes, nomenclature and database. The lack of interoperability of these custom systems results in insufficient controlling. Much stricter requirements must be met to suit legal needs.

The development and validation of not commercially available allelic ladders requires quite a large investment, but their application provides for professional genotyping. The power of discrimination and statistical evaluation strongly depends on the unique properties of the given population and the allele frequency databases. According to criminalistic experiences, the coancenstry coefficient, which expresses the genetic correlation among subpopulations, and the inbreeding coefficient, which represents the non-random mating of populations are necessary for the calculation of statistical probability. A high level of inbreeding in the same lineage, despite a relatively high power of discrimination, can result in identical DNA profiles. The amplifying effects arising from close relation and descent not only modify the legal interpretation of the power of discrimination, but also – due to the lack of the breeding kennels genetic databases – may allow incorrect parentage determination. Since no test is currently available for quantification of the low amount of canine specific target DNA extractable from heterogeneous samples, the species specific monoplex, and low level artifact producing PCR amplification is suitable for at least partial genetic profiles, and large scale screening in casework. Successful genotyping can be performed from various types of casework samples, but the success rate can be affected by the kind of traces, transfer effects and the sampling.

The results and conclusions described above provide for a well established expert opinion in connection with canine biological materials, for legal as well as breeding purposes. The category of confirmation of species identification is applicable, but risk of potential mixing or contamination requires careful consideration. The individualization can be declared in probable categories, but genotyping without validated allelic ladders during the profiling process presents a high risk to the formation of the correct expert opinion. The lack of a representative population genetic database means that correct, numeric values are not available for individualization, parentage control and

correction factors – F_{ST}, F_{IS} – cannot be incorporated. Due to the assumed highly heterogeneous nature of the Hungarian canine population, the use of concordant reference data during statistical testing is not supported by recent studies. Collaboration of the breeding association is essential for extrapolating allele frequency data in the local breed populations and breeding lines. Authoritative population surveys are necessary not only for forensic individualization, but also for professional breeders and breeder clubs to perform parentage control and screen for genetic diseases in breeding lines.

6. PUBLIKATIONS

Budowle, B., Woller, J., Koons, B.W., Füredi, S., Errera, JD., Pádár, Zs. (1996) Hungarian population data on seven PCR-based loci. J Forensic Sci 41:667-70

Egyed, B., Füredi, S., Angyal, M., Boutrand, L., Vandenberghe, A., Woller, J., Pádár, Zs. (2000) Analysis of eight STR loci in two Hungarian populations. Int J Legal Med 113:272-75

Egyed, B., Füredi, S., Angyal, M., Boutrand, L., Vandenberghe, A., Woller, J., Pádár, Zs. (2000) Analysis of eight STR loci in two Hungarian populations. Forensic Sci Int 113:25-27

Egyed, B., Csikai, M., Füredi, S., Pádár, Zs. (2005) Population genetic data on the STR loci D2S1338, D19S433 and SE33 in Hungary. J Forensic Sci 50:720-21.

Egyed, B., Füredi, S., Pádár, Zs. (2005) Population genetic study in two Transylvanian populations using forensically informative autosomal and Y-chromosomal STR markers. Forensic Sci Int Doi: 10.1016/j.forsciint.2005.10.020

Egyed, B., Füredi, S., Angyal, M., Balogh, I., Kalmár, L., Pádár, Zs. (2006) Analysis of the population heterogeneity in Hungary using fifteen forensically informative STR markers. Forensic Sci Int 158(2-3):244-9.

Füredi, S., Woller, J., Pádár, Zs. (1995) Hungarian population data for the STR systems TH01 and VWA. Int J Legal Med 108:48-49

Füredi, S., Budowle, B., Woller, J., Pádár, Zs. (1996) Hungarian population data on six STR loci - HumVWFA31, HumTH01, HumCSF1PO, HumFES/FPS, HumTPOX, and Hum HPRTB - derived using multiplex PCR amplification and manual typing. Int J Legal Med 109: 100-01

Füredi, S., Woller, J., Pádár, Zs. (1997) A population study of the STR loci HumLPL, HumF13B and HumF13A01 in Hungary. Int J Legal Med 110:107-08

Füredi, S., Angyal, M., Kozma, Zs., Sétáló, J., Woller, J., Pádár, Zs. (1997) Semi-automatic DNA profiling in a Hungarian Romany population using the STR loci HumVWFA31, HumTH01, HumTPOX and HumCSF1PO. Int J Legal Med 110:184-87

Füredi, S., Kozma, Zs., Woller, J., Pádár, Zs., Angyal, M., Bajnóczki, I., Nishi, K. (1998) Population genetic data on four STR loci in a Hungarian Romany population. Int J Legal Med 112:72-74

Füredi, S., Angyal, M., Woller, J., Pádár, Zs. (1999) Y-STR haplotyping in two Hungarian populations. Int J Legal Med 113:38-42

Füredi, S., Egyed, B., Vandenberghe, A., Angyal, M., Woller, J., Pádár Zs. (2000) Population genetic data on 5 autosomal and 8 Y-chromosomal STR loci in 2 Hungarian populations. Progress in Forensic Genetics (8) ICS 1193. Elsevier, ISBN 044450303X:151-53

Pádár, Zs., Füredi, S., Woller, J. (1997) Polimorf DNS lókuszok PCR amplifikációja. Molekuláris Medicina, (Kopper, L., Marcsek, Z., Kovalszky, I. szerk). Medicina Könyvkiadó Rt., Budapest, ISBN 963 242 345 3:112-18

Pádár, Zs., Barta, A., Egyed, B., Füredi, S., Kemény, G., Woller, J. (2000) Hungarian experience of examination of the fingernails in violent crime. Progress in Forensic Genetics (8) ICS 1193. Elsevier, ISBN 044450303X:492-94

Pádár, Zs., Füredi, S., Angyal, M. (2001) Kriminalisztikai (DNS-) vizsgálati lehetőségek újszülöttmegölésekben. Belügyi Szemle (1):69-72.

Pádár, Zs., Egyed, B., Kontadakis, K., Zöldág, L., Fekete, S. (2001) Resolution of parentage in dogs by examination of microsatellites after death of putative sire. Acta Vet Sci Hung 49(3):269-73

Pádár, Zs., Angyal, M., Egyed, B., Füredi, S., Woller, J., Zöldág, L., Fekete, S. (2001) Canine microsatellite polymorphism as a resolution of an illegal animal death case in a Hungarian Zoological Garden. Int J Legal Med 115:79-81

Pádár, Zs., Kontadakis, K., Egyed, B., Füredi, S., Woller, J., Zöldág, L., Fekete, S. (2002) Canine STR analyses in forensic practice – observation of possible mutation of a catagen doghair. Int J Legal Med 116:286-88

Pádár, Zs., Egyed, B., Kontadakis, K., Füredi, S., Woller, J., Zöldág, L., Fekete S. (2003) Importance of canine identification in the Hungarian forensic practice. Progress in Forensic Genetics (9) ICS 1239. Elsevier, ISBN 0444507175 :897-900

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

Pádár, Zs., (2004) Kriminaisztikai célú DNS-vizsgálatok Magyarországon. Kriminalisztika (Bócz, E. szerk) BM Duna Kiadó, Budapest, ISBN 963 836 83 4 ö:598-606

Pádár Zs. (2005) A DNS-vizsgálatok szerepe és szakértői problémái emberölési ügyekben. Belügyi Szemle (1):13-29

Sótonyi, P., Járay, J., Pádár, Zs., Woller, J., Füredi, S., Gál, T. (1996) Comparative study on reused haemodialysis membranes. Int J Artificial Organs 19:387-92

Woller, J., Füredi, S., Pádár, Zs. (1995) AMPFLP analysis of the VNTR loci D1S80 and ApoB in Hungary. Int J Legal Med 107:273-74

Woller, J., Füredi, S., Pádár, Zs. (1996) Hungarian population data for 11 PCR-based polymorphisms. Advances in Forensic Haemogenetics (6) Springer, ISBN 3540604928:647-49

Woller, J., Budowle, B., Füredi, S., Pádár, Zs. (1996) Hungarian population data on the loci HLA-DQα, LDLR, GYPA, HBGG, D7S8 and GC. Int J Legal Med 108:280-82

Woller, J., Füredi, S., Pádár, Zs. (1997) Polimeráz láncreakción alapuló DNS vizsgálatok a magyar igazságügyi gyakorlatban. Orvosi Hetilap 138:3223-28

Woller, J., Budowle, B., Angyal, M., Füredi, S., Pádár, Zs. (1998) Population data on the loci HLA-DQα, LDLR, GYPA, HBGG, D7S8, GC and D1S80 in a Hungarian Romany population. Progress in Forensic Genetics (7) ICS 1167. Elsevier, ISBN 0444829652:381-83

Zöldág, L., Albert, M., Fodor, Zs., Pádár, Zs., Kontadakis, K., Eszes, F. (2001) Az általános vízkór (anasarca) öröklődéskórtani vizsgálata magyarországi angol bulldog fajtájú kutyapopulációban. Hung Vet Journal (6):335-41