The Development and Validation of a Standardized Canine STR Panel for Use in Forensic Casework (NIJ Grant No. 2004-DN-BX-K007)

Sree Kanthaswamy

Abstract: The American Pet Products Manufacturers Association 2007-2008 National Pet Owners survey found that 39% of U.S. households have at least one dog (APPMA 07-08A). Given that dogs regularly come into close contact with humans, canine biological material is frequently found at crime scenes and constitues a useful source of evidence. Testimony based on short tandem repeats (STRs) analysis of animal DNA evidence has been used in a number of court cases worldwide. For example, casework based on domesticated dog DNA has been instrumental in successfully *linking a perpetrator to a crime scene in instances of murder, burglary, and sexual assault* and in cases of animal abuse and theft. Unfortunately, despite its tremendous potential for facilitating forensic investigation, canine DNA remains under-utilized as a source of evidence. The few laboratories that currently perform forensic analysis of canine DNA use a variety of marker sets with little overlap among them. Recent court challenges to canine DNA analysis have demonstrated a need for a standardized canine STR panel that has been validated according to human forensic guidelines.

In collaboration with the California Department of Justice, the National Institute of Standards and Technology, the National Institute of Health, the Federal Bureau of Investigation, MMI Genomics, and Finnzymes, we developed and validated *a canine forensic fluorescent-based multiplex composed of 18 independently-segregating STR markers that are robust, reliable, and informative for all dog breeds*. In addition to STR markers, the panel also includes the gender identification and the canine Zinc Finger gene. This multiplex was used to profile Pit Bull Terrier and Rottweiler samples (two breeds frequently involved in forensic cases); mixed-breed samples, and a collection of pedigreed pure-bred dogs (over 667 unrelated individuals representing over 50 American Kennel Club-recognized breeds). These subpopulations formed the foundation of our canine database. The resulting data was used to assess genetic and geographic substructure and to estimate recombination ratios and inbreeding coefficients, including Fst. In addition, genetic diversity among and within the regional and breed datasets were compared.

Each pedigreed dog population was genetically distinct and could be differentiated from the mixed breed dog population. Genetic diversity was slightly clinally distributed among the different U.S. regions. The results herein provide further support for using allele frequency data with the canine STR multiplex to convey the significance of identity testing for forensic casework, parentage testing, and breed assignments. The underlying scientific procedures and results have been submitted to peer-reviewed journals and presented at national and international conferences.

Table of Contents: **Page**

Executive Summary:

Routine STR profiling of human bioevidence is the mainstay of most state-sponsored and private forensic laboratories. STR analysis provides a powerful approach for match comparison and parentage or kinship testing. In human DNA testing, the DNA Advisory Board, led by peer consensus, has successfully implemented standards in quality assurance for performing DNA typing with efficiency and reproducibly (DNA Advisory Board [DAB] 1998).

Though not considered mainstream, animal forensic DNA testing is performed routinely in some private and academic laboratories to assess an animal's identity by discovering its parentage, breed, species, or geographic origin. In several instances, pet animal DNA has been used to adjudicate criminal or civil cases (Menotti-Raymond et al. 1997, Halverson and Basten 2005, Budowle et al. 2005), but forensic analysis of biological material from animals is not limited only to household pets. STR analyses of animal DNA have been presented in court cases involving livestock such as cattle, goats, horses, llamas, and sheep (http://www.vgl.ucdavis.edu/forensics/index.php). Since dogs are common domestic pets in the United States (The American Pet Products Manufacturers Association's [APPMA] National Pet Owners Survey of 2005-2006 estimated that there are over 70 million dogs in the U.S. with an average of 1.7 dogs per house-hold and about one dog for every four persons in the U.S. (2007-2008 APPMA survey)), the development and validation of a commercially available canine-specific forensic STR profiling kit would provide a valuable tool to forensic investigations.

Despite its obvious utility, domestic dog DNA evidence remains largely under-utilized in criminal investigations. While this situation is partially created by the absence of robust forensic STRs, another reason is the unavailability of standardized and validated canine PCR kits and protocols as well as a lack of standard nomenclature and internal sizing standards (allelic ladder). The lack of information about the potential value of canine hair or other biological samples and the low probability of obtaining an informative STR profile from fallen hair samples may also contribute to the indifference of the forensic community to the value of non-human DNA evidence.

Unlike human identification kits, there are no commercially available forensic STR multiplex kits for canine DNA testing despite the utility they would offer the forensic community. Zajc et al. (1994), Shutler et al. (1999), Pádár et al. (2001a, b, 2002) and Halverson and Basten (2005) have opted to use either commercial kits or laboratory-assembled STR panels originally developed for non-forensic testing in forensic applications. Commercial and laboratory kits for animal DNA typing are usually formulated for co-amplifying markers in pristine samples such as blood or buccal cells with high copy numbers of DNA. The Applied Biosystems' (Foster City, CA) Stockmarks® Canine I and II kits, once marketed for routine parentage assessment, have been used to analyze evidentiary canine material in homicide cases (Halverson and Basten 2005), but these canine kits, which neither include allelic ladders nor proposed a nomenclature for forensic casework, were discontinued from production in 2005. A vast number of forensic cases not reported in the literature have used in-house or commercially developed non-forensic kits. Most of these cases also rely on ad-hoc and private databases that were assembled based on these non-forensic reagent kits

An ideal database should include samples from a broad spectrum of sources, including pure breed dogs whose lineages are recorded by breed registries, pure breed dogs whose lineages are not recorded, dogs from pedigrees that cannot be tracked and, consequently, are not recognized by most breed registries, and dogs that are hybrids of two or more breeds. Canine population genetic databases in the U.S. were originally established via routine parentage and pedigree assessment or for other non-forensic applications and likewise are dominated by pedigreed dogs (or pure breed dogs that have their ancestry formally recorded by breed registries). Relevant databases reflecting the genetic structure of mixed breed dogs (which represent approximately 50% of the dog population in the U.S, APPMA 2005-06A) are not available. To obtain more meaningful values for calculating DNA profile frequencies, the canine

database should also be geographically representative in order to show the distribution of genetic variation across the country. Since biomaterial from dogs reflecting various types of canine heritage could potentially be present at crime scenes, our regionally representative canine database comprises pedigreed dogs, mixed breed and pure breed dogs (purebred dogs without formal pedigree records).

Until now, no concerted effort has been made by animal typing laboratories to develop a forensically validated canine STR kit and an accompanying population database following the Scientific Working Group on DNA Analytical Methods guidelines (Technical Working Group on DNA Analysis Methods [TWGDAM] 1995). Our joint initiative to assemble a canine forensic STR typing kit involved input from several private, academic, state, and federal laboratories including the Department of Justice Laboratory in Richmond, CA, MMI Genomics Inc. (MMIG) in Davis, CA, QuestGen Forensics in Davis, CA and the Molecular Anthropology Laboratory (MAL), University of California Davis (UC Davis), CA, the National Institute of Standards and Technology (NIST) in Gaithersburg, MD, the FBI DNA Laboratory, in Quantico, VA, the Laboratory of Genomic Diversity, National Cancer Institute, in Frederick, MD and the diagnostics division of Finnzymes Oy, in Espoo, Finland.

As part of a series of developmental validation measures of the proposed commercial forensic kit (Dayton 2008.), the present report provides an analysis of 18 polymorphic, canine-specific STRs that are included in the developed kit. The panel of markers, which includes the X and Y chromosome-linked zinc finger loci for gender determination, has been formatted into an easy-to-use, quality-controlled, and validated (Dayton 2008) reagent kit. This kit has been earmarked for commercialization by Finnzymes Oy and the establishment of allelic ladders based on allelic sequences and a nomenclature system (Tom et al. submitted) is underway.

We conducted population studies on three different sample sets: one included geographically representative populations of mixed breed dogs from 50 states in the U.S., a second that represented dogs from the most popular pedigreed breeds in the U.S. (according to the American Kennel Club [AKC, http://www.akc.org/]), and the third included Rottweilers and American Pit Bulls, two dog breeds which have been characterized as ‘dangerous’ by the U.S. Center for Disease Control (CDC) due to their frequent involvement in dog bite-related fatalities (Gershman et al. 1994, Sacks et al. 2000). The resulting population database contains publicly accessible information designed to be admissible in U.S. courts including data on locus informativeness, allele frequencies, distribution of domestic dog genetic variation, match probability estimates, and inbreeding coefficients. Average random match probability is 1 in $2x10^{33}$ using the regional database and 1 in $4x10^{39}$ using the breed dataset.

Here, we also report additional validation studies of the assay to determine its robustness and reliability in forensic DNA typing (Dayton 2008). These studies include species-specificity and sensitivity testing, peak height ratio and intra- and inter-locus color balance studies, mixture, annealing cycle number and temperature studies, and case type samples and population studies. The reagent kit not only robustly amplifies domesticated dog samples but also wolf (*Canis lupus*) samples. This kit is also capable of consistent amplification of full 19-locus profiles from 125 pg of canine DNA and will be commercially manufactured for the forensic research community to ensure compliance to quality assurance standards for forensic casework. Several studies, including Tom et al. (submitted) have been performed to evaluate genotype and sequence information for multiple canine STR markers (Shibuya et al. 1994; Francisco et al. 1996; Mellersh et al. 1997; Halverson et al; 1999; Neff et al. 1999; Hellmann et al. 2006) and more recently, various laboratories have multiplexed different canine STR markers for co-amplification in a PCR reaction. Earlier multiplexes usually contained from three to ten primer pairs that required multiple amplifications to type 15 or more markers (Koskinen and Bredbacka 1999; Halverson and Basten 2005; Eichmann et al. 2005; Hellmann et al. 2006). Such protocols are very similar to the human Profiler and Cofiler kits (Applied Biosystems, Foster City, CA); a sample of forensic interest would be amplified by both kits in order to type all 13 Combined

DNA Index System (CODIS) core loci. However, when forensic sample is limited, it is much more efficient as well as cost- and time-effective to amplify all polymorphic markers of interest in a single multiplex PCR reaction. This also avoids exhausting evidence which can be made available for comparative testing. Similar to the human Identifiler® kit (Applied Biosystems, Foster City, CA) designed to co-amplify all 13 CODIS loci.

The reagent kit enables co-amplification of 18 STR loci and the sex-linked Zinc Finger (for canine sex determination) marker. As a first evaluation of the proposed multiplex, developmental validation studies similar to the testing described in the Identifiler® validation paper (Meiboom et al. 2004; Collins et al. 2004) are reported in this study. The performance of the reagent kit is discussed in comparison with that of previous typing kits utilized for forensic casework, primarily Identifiler® (Collins et al. 2004), Stockmark® Canine I and II panels (Applied Biosystems, Foster City, CA [Halverson and Basten 2005]), the Eichmann et al. canine multiplexes (2004a; b), and the Menotti-Raymond et al. 'Meowplex' (2005).

In its current state, the Finnzymes Canine 2.1 Multiplex STR Reagent kit contains 18 polymorphic STR markers. The multiplex is able to type 0.125ng of DNA template in a 20uL reaction volume. The optimized PCR amplification parameters for the kit are consistent with previous canine typing systems utilized for forensic casework and are similar to feline and human typing kits. Furthermore, the Finnzymes multiplex has a level of sensitivity, ability to detect low level mixture ratios, and power of discrimination comparable to those of other forensic typing systems when amplifying case-type samples.

The proposed reagent kit is also capable of genotyping wolf (*C. lupus*) samples, which can be useful in wildlife population and conservation studies. The overall peak height ratios for most markers (excluding the Zinc Finger and FH2017) ranged between 78.88 and 93.05%. Intracolor balance was not as robust as in Identifiler® (between 32-45% depending on color channel); therefore, we suggest that the manufacturer of the canine kit improve the balance prior to the kit's commercial release. Intercolor balance was 22%, which was averaged over 61 evaluated samples. To obtain better intracolor balance, increasing the signal of markers that tended to produce the lowest RFU values per color (FH2017, PEZ05, FH2107, and FH2088) and reducing the signal in markers that tended to generate the highest RFU values (FH2010, Zinc Finger, FH2054) could be investigated.

In addition, certain primers had high sequence similarity with non-canine species. The FH2017 forward primer has 95-100% identity to numerous locations in the horse, human and rhesus monkey genomes. Also, it is probable that the primers designed for FH2017 bind to a region of DNA that contain site mutations, which is apparent in the high standard deviation for the peak height ratios calculated. The resulting high rate of allelic drop out from mutations in the FH2017 primer annealing site probably caused the lowest across-loci average observed and expected heterozygosity values obtained in a study involving 667 dogs (Kanthaswamy et al. submitted). Rather than addressing the various problems associated with marker FH2017, the primer pair should be removed from the multiplex all together. PEZ05 is another primer pair that presents problems. The PEZ05 marker is located on the same chromosome as another marker, FH2054. This usually generates the lowest signal within its dye color, and ambiguous allele calls can result from its 10 base pair overlap (discovered from the positive control animal) with FH2001. Rather than redesigning the primers, PEZ05 is another candidate for removal from the Finnzymes multiplex. Marker FH3313 is a third candidate for removal because of microvariants that are difficult to resolve due to slight sequence variations (Tom et al., submitted), and its low normalized peak signal, which reduces intracolor balance in the red channel. Even after these removals, the kit would still have 15 STR loci, which should be completely sufficient for high informativeness in forensic casework.

DNA extracted from blood and buccal cells can reliably produce full profiles when amplified with the canine STR kit. Phenol/chloroform extracted hair samples produced partial profiles. Several Chelex extracted hair samples from non-probative case type samples also

produced partial profiles, as did non-probative saliva stains on denim pants. Further studies designed to improve the typing results from low copy number samples (such as shed hair samples) should be planned. Furthermore, the preliminary differences in typing results dependant on other canine hair qualities (i.e. the coarseness of the hair) might also be a worthwhile investigation for optimizing sample profile information.

The sequence information from STR loci, on which the allelic nomenclature system is based, reflects the importance of point mutations, insertions, and deletions within and outside these loci's core repeat structures. Structural variation significantly impacts the precision of fragment sizing based on capillary electrophoresis, confounding allele calls, allele numbers, and estimates that rely on allele frequency. Because an understanding of the STRs' sequence structure will improve the accuracy and precision of allele fragment sizing in U.S. canine forensic genetic testing, the expected allele sizes have been calculated and their repeat stuctures defined based on novel allelic sequence information.

A canine STR nomenclature and allelic ladder system have been proposed to enhance accuracy in fragment sizing and to facilitate inter-lab comparison of STR data to augment STR interpretation (Eichmann et al. 2005; Hellmann et al. 2006). In accordance with the International Society of Forensic Genetics (ISFG) (DAB 1994; Bär et al. 1997; Eichmann et al. 2005; Hellmann et al. 2006), this repeat-based nomenclature system designates canine STR alleles based on the number of full repeat motifs and, separated by a decimal, the number of nucleotides contained in the partial or incomplete repeat motifs or microvariants (repeat motifs smaller in size than the general repeat motif size at a particular locus). Because this system is based on an internationally accepted nomeclature method, our goal was to use Eichmann et al.'s (2005) and Hellmann et al.'s (2006) nomenclature system to characterize the actual size and sequence structure of several canine STR loci that have not been previously characterized structurally but are included in the commercial Finnzymes multiplex.

Detailed repeat motif sequences relative to common allelic size ranges of five of the 18 STR loci (FH2010, FH2054, FH2328, PEZ02, and vWF.X) in the proposed STR reagent kit have been published and displayed on the NIST Short Tandem Repeat DNA Internet DataBase (http://www.cstl.nist.gov/strbase/dogSTRs.htm). Tom et al.'s (submitted) study has three objectives, the first of which is the verification of the STR sequences and core repeat structures of these five STRs, based on canine samples collected from domestic dogs throughout the U.S. This objective is particularly important because the primers used in our proposed STR panel do not target the same annealing sites as previously cited (Eichmann et al. 2005; Hellmann et al. 2006). The second objective of this study is to describe the sequences and core repeat units of the remaining 13 STR loci. The third objective is to characterize the sequences of the sexing marker, the canine X and Y chromomsome-linked Zinc Finger gene, that were incorporated into the panel.

Regardless of species, STRs are very informative markers for uniquely identifying an individual. The STR panel currently being developed to be incorporated into the Finnzymes Canine 2.1 reagent kit will allow biological evidence from canines to provide important information in criminal investigations. Since the loci chosen to be a part of this 19-plex are a combination of published loci and unpublished loci, it is prudent to confirm that the same repeat units are obtained for the published loci as well as to produce high quality sequences for the loci that have not yet been published.

The number of alleles present for a particular locus is an indication of the degree of polymorphism of the marker and its value for use in canine DNA testing. While sequences flanking the STR core repeat regions normally remain unchanged, sequence variation in these flanking regions also occur and can impact the amplicon sizes of each of the possible alleles, the number of alleles per locus, and estimates based on allele frequencies and interpretations of DNA profiles. Occasionally, mutations in the flanking regions can cause PCR failure and allelic drop-outs, as observed by Dayton (2008), among some of the canine loci studied here.

This study adopted the nomenclature system described by Eichmann et al. (2005) and Hellmann et al. (2006) by using the first detectable repeat unit as a starting point to characterize the repeat region and the last repeat as the ending point. It would be detrimental to base STR nomenclature on the specific variable locations within the repeat regions and not the entire repeat region because as more samples are analyzed using these STR loci, new (rare) alleles will be discovered and need to be included in the database without confounding the existing alleles at a specific locus.

An important component for the proposed reagent kit is the development of internal sizing standards (i.e., allelic ladder) for standardizing allelic designations (Hellmann et al. 2006). The allelic ladder should contain the common alleles with basic repeat motifs and common microvariants of the STR loci comprising the reagent kit for use as reference for fragment sizing. The sequences of alleles used as rungs in the allelic ladder should be established so that size determination of unknown alleles can be reliably based on the similarity of electrophoretic mobility of known alleles in the ladder. The known sequences of the alleles in the allelic ladder will also facilitate inter-laboratory comparisons even if different samples are used to develop the allelic ladders.

The human STR nomenclature is based on the 1994 recommendations of the International Society of Forensic Genetics (ISFG). There were no standardized canine STR nomenclature and allelic ladder systems in existence until Eichmann et al. (2005) and Hellman et al. (2006) proposed a method system that has yet to be adopted by other laboratories in canine forensic STR genotyping. Standardized and validated animal STR protocols that emerged from these previous studies as well as the one study currently being completed will promote the acceptance of animal DNA typing through creation of validated STR panels, publication of peer-reviewed articles, and establishment of a centralized database. The use of a standardized panel of loci in all investigations would facilitate interlaboratory comparison of information and will enhance the accuracy and precision of canine forensic genetic testing in the U.S. The allelic nomenclature proposed here in conjuction with a breedwise and regional-wide population genetic study using the markers included in the reagent kit by Kanthaswamy and colleagues (submitted) and the developmental validation of the kit by Dayton (2008) will provide a more accurate description of STR allele frequency distributions among the various U.S. canine subpopulations.

This study focused on the application of the Finnzymes multiplex and associated genetic database in the forensic genetic identity and parentage testing of canines in the U.S. The kit's panel of 18 STRs was shown to be reproducible, informative, and robust. The database, which is constructed based on the STRs' allele diversity and frequency distributions, is comprehensive in terms of regional representation of pedigreed and mixed breed dog populations in the U.S. The genetic profiles and allele frequencies of important dog breeds in the U.S. that are popular as house pets and/or dangerous as vicious animals linked to fatal dog bites are also represented in the database. With their enhanced informativity and efficiency and their easy accessibility to forensic laboratories, the kit and the accompanying population genetic database combine to form a consolidated and valuable resource that could potentially develop into a universally accepted canine forensic STR system.

While the aforementioned objectives of developing a validated multiplex STR panel for forensic profiling of canine evidence, complete with a nomenclature system for enhanced reproducibility and precision, were the direct goals of the project that was proposed to NIJ, the following are additional achievements of the NIJ funded project that include the development of a real time qPCR technique for canine forensic DNA and the establishment of locality-specific and national canine mtDNA databases.

Using quantification data to control the amount of template DNA in the polymerase chain reaction (PCR), forensic scientists can optimize testing and minimize the consumption of limited samples. The ability to identify and quantify trace DNA in mixed-species samples is crucial when it may be overwhelmed by nontarget DNA, as in cases of dog attack. We evaluated two

quantitative real-time PCR assays for dynamic range, species specificity, and inhibition by humic acid (Evans et al. 2007). While both assays proved to be highly sensitive and discriminating, the Melanocortin-1 Receptor (MC1R) gene Taqmans assay had the advantages of a shorter run time, greater efficiency, and safer reagents. In its application to forensic casework, the MC1R assay has been advantageous for quantifying dog DNA in a variety of mixed-species samples and facilitating the successful profiling of individual dogs.

Mitochondrial DNA (mtDNA) analysis of hypervariable 1 (HV1) region in domestic dogs was carried out to determine the power of HV1 polymorphic sites for identifying individual dogs and breed type. A 608 bp-long stretch of canine mtDNA comprising the HVI region was sequenced for 36 domestic dogs with varying degrees of breed admixture representing 19 different distinct breeds. Evaluation of the haplotype sequence variation revealed sixteen unique haplotypes with frequencies ranging from 3 to17 percent, confirming that the exclusion capacity of mtDNA is significantly lower than nuclear DNA markers, particularly a panel of informative STRs.

No correlation was observed between mtDNA haplotype and type or breed of dogs based on morphological features or the owner's verification. However, some haplotypes were more disproportionately represented in some breed mixes than other haplotypes. Because breed types were first verified by the owners of the dogs, the results of this study suggest that testimony by witnesses regarding the breed of a dog involved in a forensic investigation is not reliable.

A 60 bp "hotspot" within the canine mtDNA HVI was discovered that portends a valuable molecular tool for future canine forensic application, particularly for assaying degraded DNA samples. The analysis of this 60 bp variation hotspot sequence in the canine mitochondrial DNA (mtDNA) hypervariable region I (HVI) was conducted to evaluate its utility in forensic investigations. Sixteen haplotypes containing 15 single nucleotide polymorphisms (SNPs) were observed among 118 sequences from five regional localities in the U.S. Another nine haplotypes containing seven SNPs were detected when 177 GenBank sequences were included in the study. Assays using these haplotypes were robust, canid specific, and are expected to provide a rapid method for correctly excluding individual dogs as non-contributors. Samples from the different localities were highly variable and representative of the much broader collection of geographically representative GenBank sequences. As previous canine mtDNA studies have shown, the inclusion of the GenBank sequences in this study did not significantly increase exclusion capacity estimates suggesting that the choice of a representative database for ascertaining frequencies and exclusion probabilities is important. In both studies, we showed that the inclusion of more globally representative mtDNA sequences does not significantly increase the exclusion capacity estimates, suggesting that the choice of a representative database used to obtain frequency estimates for a particular DNA sequence can impact exclusion probabilities. Each local sample was shown to be highly variable and representative of the much broader geographic sample, reiterating the findings of Himmelberger et al. (2008).

Finally a study including the 608 bp sequence within the hypervariable region 1 (HV1) of canine mitochondrial DNA (mtDNA), first reported by Himmelberger et al. (2008), was further examined in samples of mix breed dogs within the western, midwestern, northeastern, and southern regions of the United States in order to remove biases in the current pure breed mtDNA databases (Smalling et al. in prep.). Forty-six haplotypes were discovered including the 16 haplotypes Himmelberger et al. (2008) previously reported. Analysis of molecular variance (AMOVA) concluded that all variation within the regional groups is highly correlated with geographic transitions, i.e. there is significant geographic subdivision among the four regional samplesets. These results are in stark contrast to the results that were obtained from an AMOVA analysis of short tandem repeats (STRs). The domestic dog mtDNA showed sharp geographic differences but no differences among breeds whereas STR-based data showed discrete breed-wise differences while regional differences were insignificant. Contrary to the previous findings based on mtDNA including Himmelberger et al. (2008) and Baute et al. (2008), the study by

Smalling et al. (in prep) has demonstrated the ***need*** for more locality-specific canine mtDNA databases for correctly interpreting the meaning of a mtDNA haplotype match at the local level.

This NIJ grant has sponsored the research of six Masters of Forensic Science graduate students at the University of California, Davis.

This final report herein will only focus on the development and validation of the Canine 2.1 STR Multiplex reagent kit and the direct goals of the project that was proposed to NIJ.

Main Body of the Final Technical Report

I. Introduction:

While STR analysis of human biological evidence is widely accepted by the criminal justice system, the potential for widespread acceptance of animal DNA evidence was diminished by the September 2003 ruling by the Washington Court of Appeals which excluded canine DNA evidence in the 1998 case of the State of Washington V. Kenneth Leuluaialii. The appellate court ruled that the trial court failed to establish "*whether the scientific community generally accepted that the specific loci used by PE Zoogen in the present case were highly polymorphic and appropriate for forensic use.*" The court also states that "*current canine DNA testing and mapping focuses on the goals of paternity testing, breed testing, and cancer and research studies. There is little indication that polymorphic loci and alleles in canine DNA have been sufficiently studied such that probability estimates are appropriate for the forensic use applied in this case"* (Docket Number 43507-8-I). As a consequence of this ruling, canine DNA evidence has been excluded from pending court cases in California and is threatened to face a similar fate in other states. Clearly there is a need for a validated canine STR panel that meets the rules of scientific acceptance and reliability.

Commercially available forensic STR multiplex kits have have facilitated the ease, efficiency, and standardization of human STR typing, however there are no such kits currently available for canine DNA testing. Zajc et al. (1994), Shutler et al. (1999), Pádár et al. (2001a, b) and Halverson and Basten (2005) have circumvented this problem by using in-house assembled STR panels as well as kits intended for routine parentage testing. These commercial and in-house kits for animal DNA typing were originally designed for analyzing pristine samples such as blood or buccal cells which typically contain high quality and quantity DNA. Prior to 2005, the Stockmarks® Canine I and II kits (Applied Biosystems, Foster City, CA) were used effectively to analyze evidentiary canine material in homicide cases (2). However, these canine kits did not include allelic ladders and required a level of familiarity and scientific expertise incompatible with forensic laboratories. Furthermore, the formalized nomenclature for the Stockmarks® loci was never published, and the kits were never updated to the five-dye systems now commonly used in the forensic community. Although the Stockmarks® Canine kits were discontinued from production in 2005, six of the loci in the multiplex described herein were included in the Stockmarks® kits.

Since their domestication in East Asia nearly 15,000 years ago (Savolainen et al. 2002), dogs (*Canis familiaris*) have been selectively bred to be outdoor working animals as well as household companions (Brewer et al. 2002). An estimated seventy-two million pet dogs currently live in the U.S.A. (U.S. Pet Ownership and Demographic Sourcebook 2007), and because many dog owners live in close proximity to their pets, canine DNA evidence may often be associated with crimes. Shed dog hairs can be transferred among individuals involved in a crime and can link a suspect to a crime scene, a suspect to a victim, or a victim to a crime scene (Shutler et al. 1999; Halverson and Basten 2005). In addition, biological evidence from dogs have been be used in dog attacks and abuse cases (Eichmann et al 2004a).

Several studies have generated genotype and sequence information for multiple canine STR markers (Shibuya et al. 1994; Francisco et al. 1996; Mellersh et al. 1997; Halverson et al; 1999; Neff et al. 1999; Hellmann et al. 2006), and more recently, various laboratories have multiplexed different canine STR markers for co-amplification in a single PCR. Initially multiplexes usually contained from three to ten primer pairs that required multiple amplifications to type 15 or more markers (Koskinen and Bredbacka 1999; Halverson and Basten 2005; Eichmann et al. 2005; Hellmann et al. 2006). Such protocols are very similar to that of the human DNA analyses Profiler and Cofiler kits (Applied Biosystems, Foster City, CA) in which a sample of forensic interest is amplified by both kits in order to type all 13 Combined DNA Index System (CODIS) core loci. However, when a forensic sample is limited in quantity, it is much more efficient as

well as cost- and time-effective to amplify all polymorphic markers of interest in a single multiplex PCR. Indeed a single multiplex with similar sensitivity to fewer loci multiplexes reduces evidence consumption, leaving sample for comparative testing if desired.

The available canine population genetic databases in the U.S. were originally established via routine parentage and pedigree assessment or for other non-forensic objectives and are predominated by pedigreed dogs (or pure breed dogs that have their ancestry formally recorded by breed registries). Although mixed breed dogs represent approximately 50% of the dog population in the U.S. (APPMA 2005-06A), Halverson and Basten (2005) only included 69 mixed breed dogs out of a total of 558 dogs in their database. In addition to a larger sampling of mixed breed dogs, an ideal canine database would include geographic sampling to assess the distribution of genetic variation across the country.

Goals and Objectives

1. To assemble a standardized and validated canine forensics panel by selecting the most suitable markers from existing canine forensic panels and,
2. To establish a database of canine STR genotypes that includes a validated population genetics information content to strengthen the validity of canine evidentiary data during courtroom trials. (This database will be designed in an updatable format so as to allow the uploading of new data submitted by other labs that use our multiplex panel).

In addition to these original goals, we introduced additional goals part-way through the project to ensure that our research would result in a panel of markers that was not only evaluated and validated but a reagent kit that could be easily introduced into existing crime labs. The additional goals included the following:

a. To include more samples of pure breed, mixed breed, and out-bred dogs from the different parts of the U.S. – this is to ensure better regional representation in our final database. Additional purebred dog samples from the American Kennel Club (AKC) will also be used in the project.
b. To identify and sequence the most common allele of each informative STR locus in order to establish the repeat motifs and allele assignment.
c. To perform a population study on all candidate STRs, in which the alleles will be identified and grouped according to their estimates fragment length using fixed allelic bins. The allelic size will be confirmed using sequencing analysis of length homozygotes of the most common alleles.
d. To propose a nomenclature that refers to the internationally recognized recommendations for human-specific STR loci for forensics applications (Butler 2005).
e. To evaluate the precision of the fragment size estimation on a capillary electrophoresis platform and demonstrate reproducibility of fragment length estimation for single base-pair intermediate alleles.
f. Prepare a sample bank including a 3 generation pedigree that will be made available to forensics laboratories for purposes of record exchange and comparison or proficiency tests.
g. Develop and commercialize a kit containing pertinent PCR reagents and an allelic ladder system for each locus in the final multiplex that would facilitate the introduction of the standardized panel to new laboratories. Since the establishment of a robust allelic ladder system may be cost and time-prohibitive, it may require two phases of input; a preliminary one under the current grant that will include cloning and sequencing alleles from the final selection of markers followed by the second phase which will deliver the final product, hopefully with a new grant specifically funded for the development of the system.

Based on the markers that are already available in the animal forensics community, a standardized panel entailing highly informative and robust markers can be achieved through the

cooperation from the Department of Justice Laboratory in Richmond, CA, MMI Genomics Inc. (MMIG) in Davis, CA, QuestGen Forensics in Davis, CA and the Molecular Anthropology Laboratory (MAL), UC Davis in Davis, CA, the National Institute of Standards and Technology (NIST) in Gaithersburg, MD, the FBI DNA Laboratory, in Quantico, VA, the Laboratory of Genomic Diversity, National Cancer Institute, in Frederick, MD and the diagnostics division of Finnzymes Oy, in Espoo, Finland. As noted above, each laboratory's panel has certain loci that uniquely meet the needs of that laboratory, but they also have a proportion of markers in common. Collectively, these markers (especially those that overlap) could potentially provide a starting point to form a consensus panel that is derived from the various panels used by each laboratory.

The crux of our proposed approach is to compare and select STR markers from previously established canine panels (such as the SuperPlex, the StockMarks and other panels that are primarily used by the canine testing community) and consolidate a standardized panel with 20 unlinked tetranucludotide loci.

The markers we select must exhibit most of the following critical genetic and technical criteria that ensure optimal and reliable information underpinning aspects of breed, regional, and identity testing and significantly improve the precision and efficacy in high throughput genotyping (Kanthaswamy et al. 2006):

1. Absence of linkage;
2. High probability of parentage exclusion and individual genetic identity;
3. High level of gene diversity and polymorphic information content;
4. Presence of some alleles informative of regional origin within the U.S;
5. Tetranucleotide repeat motif;
6. Low occurrences of mutations and null (undetected) alleles;
7. Low intra-locus sampling error for estimating baseline genetic parameters (such as gene diversity and population substructure);
8. Accuracy of allele characterization (e.g., due to absence of stutter bands, non-specific amplification, etc.);
9. Balance between the length of PCR products and the ability to form a successful multiplex reaction.

A multi-laboratory joint initiative has resulted in selection of 18 STR loci for development of a commercial canine DNA profiling kit, the Finnzymes Canine 2.1 STR Multiplex Reagent Kit, for forensic applications. Similar to the human Identifiler® kit (Applied Biosystems, Foster City, CA) designed to co-amplify 15 human STR loci and the amelogenin locus, the canine beta-version STR reagent kit was developed to enable amplification of 18 STR loci and the canine sex-linked Zinc Finger marker (Meiboom et al. 2004; Collins et al. 2004).

As a first evaluation of the proposed multiplex kit, similar developmental validation studies were carried out as described for validation of the Identifiler® kit (Collins et al. 2004). The performance of the reagent kit is discussed in comparison with that of previous typing kits utilized for forensic casework, primarily Identifiler® (Collins et al. 2004), Stockmark® Canine I and II panels (Applied Biosystems, Foster City, CA [Halverson and Basten 2005]), the Eichmann et al. canine multiplexes (Eichmann et al. 2004a;b; 2005) and the 'Meowplex' (Menotti-Raymond et al. 2005).

Population studies were conducted on a sample set that included geographically distributed populations of mixed breed dogs from the U.S., a sample set that represented dogs from the most popular pedigreed breeds in the U.S. (according to the American Kennel Club [AKC, http://www.akc.org/]), and sample sets of Rottweilers and American Pit Bulls, two dog breeds which have been characterized as 'dangerous' by the U.S. Center for Disease Control

(CDC) because of the frequency of their involvement in dog bite-related fatalities (Gershman et al. 1994; Sacks et al. 1998). In addition to their importance in mauling incidences, the inclusion of Pit Bulls in this study is of further significance because Pit Bulls are the preferred breed in dog-fighting circles and are thought to be more out bred due to frequent interbreeding with breed types that exhibit the desired phenotypes of aggressiveness and morphology. As such, both the Pit Bull and Rottweiler sample sets represent excellent models for determining the efficacy of our proposed markers in breed and individual identification.

II. Methods

a) Population genetics study and genotype database development (Kanthaswamy et al., submitted).

DNAs used in this study are from MMIG's collection which includes samples from pedigreed dogs registered with the United Kennel Club (UKC, http://www.ukcdogs.com/WebSite.nsf/WebPages/Home) and were extracted from cheek swabs using methods described by DeNise et al. (2004).

The PCR was performed in 20 µL volumes containing 2 µl (1.0 ng/µl) template DNA, 9 µl Reaction Mix including Phusion™ Hot Start High-Fidelity DNA Polymerase (Finnzymes Oy, Keilaranta, Finland) and 9 µl Primer Mix using the AB GeneAmp® PCR System 9700® PCR System (Applied Biosystems, Foster City, CA). The thermal cycling parameters were 98ºC for 3 min; then 30 cycles at 98ºC for 15s; 60ºC for 75s and 72ºC for 30s followed by a final 72ºC for 5 min. For allele typing, 1.5 µl of the amplified product and 0.3 µl of GeneScan-500 [LIZ]® Size Standard (Applied Biosystems, Foster City, CA) were added to 10 µl of deionized formamide (Amresco, Solon, OH), denatured at 95ºC for 3 min and snap cooled for 3 min on crushed ice. Post-PCR amplification products were diluted 1:30 (DNA: high purity water). Two microliters of the diluted amplified product and 0.15 µl of GeneScan-500 [LIZ]® Size Standard (Applied Biosystems, Foster City, CA) were added to 10 µl of Hi-Di™ formamide (Applied Biosystems, Foster City, CA), denatured at 95ºC for 3 min and followed by snap cooling for 3 min on a StrataCooler Benchtop Cooler (Stratagene, La Jolla, CA).

Electrophoresis was conducted on an ABI PRISM™ 3130-*Avant* Genetic Analyzer (Applied Biosystems, Foster City, CA) or on an ABI PRISM™ 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA) in accordance with the instructions in the Finnzymes' Canine Genotypes™ Panel 2.1 manual. Using the Fragment Analysis 36_Pop4 module, the PCR products were injected for 10s at 3.0 kV, and subjected to electrophoresis at 15.0 kV at 60ºC using the Performance Optimized Polymer (POP™ 7 or POP™ 4, Applied Biosystems, Foster City, CA) in a 36 cm capillary. Fragment sizing was conducted with comparison to a positive control (the Canine Genotypes™ Control DNA001). GeneMapper® Software v4.0 collection and analysis, ABI PRISM® Data Collection Software v1.1 and GeneScan Analysis v3.7 software packages were used for data collection and size estimation of the fluorescent labeled DNA fragments. ABI PRISM® Genotyper v3.7 NT software was used for automated genotyping of the samples.

Since an allelic ladder for internal sizing was still under development at the time of the study (Tom et al. submitted), the PCR products were binned into their respective allelic categories using the FLEXIBIN program and methods described by Amos et al. (2007). The program was modified by the authors to accommodate the repeat motifs of the FH3313 and vWF.X loci (W. Amos, Pers. Comm.). The entire data set of raw and binned alleles used in this study is available at http://www.cstl.nist.gov/biotech/strbase/.

The map coordinates and other relevant information about the 18 autosomal STRs and the sex-linked zinc-finger markers are listed in Table 1. With the exception of VWF.X, a hexameric marker and FH3377, a pentameric marker, all STRs are tetrameric. These include four

pairs of syntenic markers: FH2107 and FH3377 on chromosome 3, FH2054 and PEZ05 on chromosome 12, FH2017 and FH2088 on chromosome 15 and PEZ 16 and vWF.X on chromosome 27.

Only samples with genotypes for all loci were used in our population genetic analysis. The pedigreed dog sample of 236 animals represented nine officially recognized dog breeds, including American Pit Bull Terrier (N = 38), Beagle (N = 34), Dachshund (N = 3), German Shepherd (N = 35), Golden Retriever (N = 32), Labrador Retriever (N = 38), Poodle (Miniature Poodle [N = 15],, Toy Poodle [N = 12], and Standard Poodle [N = 8]), Rottweiler (N = 15), Shih Tzu (N = 4) and Yorkshire Terrier (N = 2). In this study, the Poodles were separated according to their varieties and treated as three separate breeds; therefore, twelve separate pedigreed breeds were actually analyzed. The 431 mixed breed dogs used in this study represent various combinations of 43 different breeds (including Afghan Hound, Akita, American Pit Bulls, Basenji, Basset Hound, Beagle, Belgian Tervuren, Bernese Mountain Dog, Border Collie, Borzoi, Boxer, Bulldog, Chihuahua, Chinese Shar Pei, Chow Chow, Cocker Spaniel, Collie, Dachshund, Doberman Pinscher, English Setter, German Shepherd Dog, German Shorthaired Pointer, Golden Retriever, Greyhound, Italian Greyhound, Labrador Retriever, Mastiff, Miniature Pinscher, Miniature Schnauzer, Mongrel, Poodle, Pug, Rottweiler, Saluki, Samoyed, Scottish Terrier, Shetland Sheepdog, Shih Tzu, Siberian Husky, St Bernard, Staffordshire Bull Terrier, Whippet, and Yorkshire Terrier) as determined by the Canine Heritage Breed Test ^{TM}. The samples of pedigreed and mixed breed dogs were acquired from across the U.S. and subdivided into the western (N = 147), southern (N = 241), mid-western (N = 164) and northeastern (N = 115) regions for part of the analysis (U.S. Census Bureau [http://www.census.gov/field/www/], see map in Figure 1). We have also prepared a sample bank including a three generation pedigree that will be made available to forensics laboratories for purposes of record exchange and comparison or proficiency tests (Kou and Kanthaswamy, in prep.).

The exact probability test in GENEPOP version 3.4 (Raymond and Rousset 1995) was used to test for the presence of linkage disequilibrium (LD, the non-random association of genotypes occurring at different loci) between the pairs of the 18 STR loci. To test the null hypothesis that genotypes at one locus segregate independently of genotypes at any other locus at the 0.05 and 0.01 levels of probability, unbiased estimates were made through randomization (1000 iterations). The Markov-chain method was used to create a contingency table representing the random association of genotypes at all possible pairs of loci. Fisher's method and a sequential Bonferroni-type procedure were used to correct for multiple significance tests (Rice 1989). In addition to examining LD between loci, Hardy-Weinberg (HW) equilibrium within each locus was analyzed using the GENEPOP software program. Allele frequencies (observed and expected), heterozygosities, hierarchical F-statistics, and pairwise Fst (Weir and Cocekerham 1984) were computed based on data from all loci using GENEPOP.

To determine if the pedigreed and mixed breed dogs' nuclear genetic variation at the 18 STRs followed a geographic pattern, the STRUCTURE 2.1 software program (Pritchard et al. 2000; Faluch et al. 2003) was used to calculate the expected allele frequencies of individual dogs in each geographic region based on an assignment index and to determine the relative probabilities of assigning each dog to each of the four regions based on the animal's genotypes. Similarly, to distinguish the allele frequencies of the pedigreed and mixed breed populations, STRUCTURE was used to probabilistically assign each dog to a breed category. Both analyses were conducted assuming an admixture model (where animals can represent a mixture of two or more ancestral groups) and correlated allele frequencies among regions and among breeds, respectively. Therefore, when a genotype reflects an admixture, or the absence of genetic substructure, a dog will be assigned to two or more populations with probability Q, the proportion of its genome that originated from the K^{th} population (Flush et al. 2003). K-values of two to four regions and two to thirteen breeds/types (representing mixed breed dogs and the three

Poodle types - the Miniature, Toy and Standard), respectively, were tested so as to include all numbers of possible populations. All STRUCTURE analyses were run at sweeps of 10^4 iterations after a burn-in period of 10^4 with and without *a priori* population information.

The accuracy of assigning individuals to their breed of origin based on genotype data was studied using individual assignment tests, which were implemented in the program GeneClass v.2.0g (Piry et al. 2004). The program includes several assignment methods, but only the Bayesian statistical approach (Piry et al. 2004) was applied due to its known efficacy (Cornuet et al. 2000).

Principal component analyses (PCAs) on the regional and breed data sets were also performed using the adegenet 1.1 package for R (Jombart 2008).

b) Development validation studies (Dayton 2008)

The Multiplex Kit and Characterization of genetic markers

The canine reagent kit contains 19 primer pairs and associated reagents to amplify 18 STR loci and one gender determination locus. Table 1 presents a list of these markers and information on repeat motifs, size ranges, map location, and the fluorescent dye tags attached to the 5' end of the forward primers, as they appear during laser excitation using filter set G5 (Applied Biosystems, Foster City, CA). A few markers are located on the same chromosome: FH3377 and FH2107 loci on chromosome 3; FH2054 and PEZ05 loci on chromosome 12; FH2017 and FH2088 loci on chromosome 15; and PEZ16 and vWF.X loci on chromosome 27. A separate study by Kanthaswamy et al. (submitted) has demonstrated that alleles of these physically linked loci segregate independently, i.e., there was no detectable linkage disequilibrium for these syntenic marker pairs. Figure 2 presents the complete profile of the positive control sample F-863 and the range of allele distribution at each locus.

Sample collection and extraction

Three different types of biological samples were typed to evaluate the 19-plex reagent kit. Liquid blood samples were kindly provided by a West Sacramento veterinary diagnostic center. Buccal and hair samples were collected from dogs throughout the state of California, primarily from the Los Angeles area, the San Francisco Bay area and the Sacramento area. Plucked and shed hair samples were collected from seven dogs that had also provided buccal samples.

DNA from liquid blood samples was extracted using the Qiagen BloodMini kit following the kit's liquid blood extraction protocol (Qiagen, Hilden, Germany). DNA from buccal samples was extracted using the Epicentre® Catch-All™ (Epicentre, Madison, WI) swabs according to the manufacturer's recommendations. Hair samples from six of the dogs were extracted using Epicentre QuickExtract™ DNA Extraction solution buffer (Epicentre, Madison, WI) according to the manufacturer's instructions. Hair samples from six dogs were also extracted using the California Department of Justice Organic Extraction protocol, with the exception that an Amicon®uLtra-4 (Millipore, Billerica, MA) centrifugal filter was used instead of a Centricon®-100 (Millipore, Billerica, MA) filter. Five of the seven dog samples were extracted using both methods.

Pre-extracted and quantified samples and standards

Non-canine DNA samples from various species were provided by the Jan Bashinski DNA Labaratory, Method Development Group of the California Department of Justice, Richmond, CA. DNA was extracted from cat, pig, horse, cow and chicken whole blood and quantified by an electrophoretic method in a 1% agarose gel containing ethidium bromide. Other previously quantified genomic DNA samples provided by the Jan Bashinski laboratory included: fish (Catfish), monkey (African green monkey), rat (Zyagen Labs, San Diego, CA), mouse (Promega, Fitchburg, WI), *Bacillus subtilis*, *Staphylococcus epidermidis*, *Candida albicans* (ATCC,

Manassas, VA), *Escherichia coli*, and *Clostridium perfringens* (Sigma, St. Louis, MO [Hudlow et al. 2008]). A human buccal sample was extracted using the Epicentre QuickExtract™ protocol (Epicentre, Madison, WI). 5 ng of DNA from each species' sample were amplified in duplicate using the canine STR reagent kit. The F-863 canine positive DNA control originating from a female Cocker Spaniel was provided in the kit. Although stated on the tube that the canine control was a 1ng/uL concentration, quantification with PicoGreen® (Invitrogen, Carlsbad, CA) indicated that the control was in fact 0.1ng/uL concentration.

Case samples

Seventeen samples from a dog bite case were used for this study. Thirteen of the samples were single shed hairs found on the clothing of the victim. All shed hairs were extracted using the Chelex extraction method. Four of the samples were saliva stains extracted from cuttings around tooth marks found in the victim's denim jeans. All saliva stains were extracted using Qiagen spin columns (Qiagen, Hilden, Germany).

Sample Quantification

All DNA samples were quantified using Quant-iT™ PicoGreen® (Invitrogen). Prior to quantification, a solution was prepared containing 15uL of PicoGreen® and 15mL TE^{-4} buffer. 100uL of Picogreen® solution were added to dilutions of 4uL DNA sample and 96uL of sterile water in a fluoroplate. K562 human DNA standards (Invitrogen) for 5ng/uL, 2.5ng/uL, 1.5ng/uL, 0.75ng/uL, 0.25ng/uL, and 0.0ng/uL were also pipetted in aliquots of 4uL into 96uL of sterile water and combined with 100uL of Picogreen solution in the fluoroplate. A Fluoroskan Ascent (Thermo Scientific, Waltham, MA) fluorometer was used to detect and compare the fluorescence values of the known standards and those of the samples to determine DNA concentrations.

Amplification

All PCR amplifications were carried out using the canine STR reagent kit. Each reaction contained 9uL of Primer Mix, 9uL of Master Mix (containing a modified version of the Phusion™ Hot Start DNA Polymerase, Finnzymes Oy, Finland), and 2uL of DNA template. PCR conditions for the standard sample study, reproducibility study, sensitivity study, mixture study, peak height ratio/stutter percentage study, species specificity study, and population study were 3 min initial denaturation step at 98ºC followed by 30 cycles of denaturation at 98ºC for 15s, primer annealing at 60ºC for 75s, and extension at 72ºC for 30s. The final step was a 5 min long extension at 72ºC, which was recommended by the manufacturer. The amplifications for these studies were performed on a GeneAmp® 9700 PCR thermocycler (Applied Biosystems, Foster City, CA). The Phusion polymerase makes use of double-stranded DNA binding domain (Sso7d) that is covalently linked to a *Pyrococcus*-like DNA polymerase domain (Wang et al. 2004). The DNA binding domain increases the processivity of the DNA polymerase by approximately 10-fold (Wang et al. 2004). Unlike the *Thermus aquaticus* (Taq) polymerase which is applied by all other STR typing kits, the Phusion polymerase also has 3' to 5' exonuclease activity (proofreading activity) that eliminates non-templated nucleotide additions during the PCR cycling process and makes the long final extension period redundant (Nord et al. 1997; Wikman et al. 2004; Phusion Hotstart Manual).

To assess the annealing temperature range of the canine 19-plex, two blood-extracted samples and one buccal-extracted sample were amplified on an Eppendorf Mastercycler epigradient thermocycler (Eppendorf, Hamburg, Germany), which creates an annealing temperature gradient across the plate head. The exact annealing temperatures tested with all three samples (in duplicate) were 56ºC, 57.7ºC, 58.5ºC, 60°C, and 61.9ºC. The Eppendorf Mastercycler thermocycler (Eppendorf, Hamburg, Germany) was also used in the reproducibility study using a 60°C annealing temperature.

Electrophoresis, Detection and Analysis

All PCR products were separated by capillary electrophoresis on an Applied Biosystems 3130 Genetic Analyzer using 36cm capillary filled with POP-4™ polymer (Applied Biosystems, Foster City, CA). Unless otherwise stated, 1 µL of PCR product was combined with 0.21 µL GeneScan™-500 LIZ® Size Standard (Applied Biosystems, Foster City, CA) and 9.79 µL Hi-Di™ Formamide (Applied Biosystems, Foster City, CA) and pipetted into an injection plate. The plate was covered with Genetic Analyzer septa (Applied Biosystems, Foster City, CA), and the contents thoroughly mixed and spun down. Samples were denatured at 95ºC for 3 min then snap-cooled on a cold block for 3 min. A 5s electrokinetic sample injection was performed with an injection voltage of 3 kV followed by a run voltage of 15 kV. The data collection software was set to detect G5 dye chemistry. After collection, the data were analyzed with GeneMapper™ ID Version v3.2.1 Software (Applied Biosystems, Foster City, CA) using 50 RFU peak amplitude detection thresholds for all colors, excluding LIZ peaks which were set to a threshold of 100 RFU.

Balance Calculations

Three different types of peak balance were evaluated to assess the overall Finnzymes canine STR kit performance (Collns et al. 2004) and to identify possible primer pairs that amplify poorly in relation to the rest of the multiplex. Intracolor balance, intercolor balance, and heterozygous peak height ratios (PHR) were calculated. Intracolor balance was calculated to assess the overall color balance within a dye lane. First, homozygous and heterozygous peaks within a dye lane had to be normalized. This was done by averaging the RFU values for each allele of heterozygous genotypes of polymorphic loci and by dividing the RFU of each homozygous peak RFU by two. The lowest normalized RFU value was then divided by the highest normalized RFU value and expressed as a percentage. Intercolor balance was calculated by first normalizing all of the loci peaks within a sample. Then the lowest normalized RFU in the sample was divided by the highest normalized RFU in the sample (regardless of dye color) (Collns et al. 2004). PHR were calculated for heterozygous genotypes for all markers by comparing the RFU values for each allele and dividing the lower RFU value by the higher value. The ratio was multiplied by 100 to express PHR as a percentage.

c) Nomenclature development (Tom et al., submitted)

Sample Selection:

The samples used in this study were collected by MMI Genomics Inc. (MMIG) in Davis, CA and extracted from cheek swabs using methods described by DeNise et al. (2004). The samples include those from pedigreed dogs registered with the United Kennel Club (UKC, http://www.ukcdogs.com/WebSite.nsf/WebPages/Home). The panel is composed of 18 autosomal STRs and the sex-linked zinc-finger locus. With the exception of VWF.X, a hexameric marker and FH3377, a pentameric marker, the remaining STRs are tetrameric. These include four pairs of syntenic markers: FH2107 and FH3377 on chromosome 3, FH2054 and PEZ05 on chromosome 12, FH2017 and FH2088 on chromosome 15, and PEZ 16 and vWF.X on chromosome 27.

The actual size and sequence structure of STR alleles were characterized by sequencing and following the recommendations of Eichmann et al. (2005) and Hellmann et al. (2006). Alleles that occurred at frequencies ≥5% in the entire sample set of 667 mixed and pure bred dogs were selected for sequencing. GENEPOP (version 3.4) was used to estimate allele frequencies (Raymond and Rousset 1995). Since a canine allelic ladder for internal sizing is not currently available, all amplicons were binned into their respective allelic categories using the FLEXIBIN program and method described by Amos et al. (2007). The program converted the observed allele sizes into the effective length of the repeat, allowing the selection of individual

dogs that carried a certain bin and allele size of interest. A representative set of the relatively few variants of intermediate size were also sequenced to determine the nature of their sequence structure.

Polymerase Chain Reaction (PCR) Amplification

DNA samples were added to a PCR cocktail containing 2 µL 10x PCR Buffer, 0.6 µL $MgCl_2$ (50mM), 0.45 µL dNTPs (10mM), 0.6 µL forward primer (10µM), 0.6 µL reverse primer (10µM), 0.2 µL Platinum Taq Polymerase (Invitrogen, Carlsbad, CA), and 9.95 µL dH_2O. Samples were amplified using a MJ Research Inc. PTC 100 Programmable Thermal Controller (GMI Inc., Ramsey, MN). All parameters except the annealing temperatures, which are provided in Table 2, were kept constant for all of the samples. PCR conditions were 95° C for 12 min., followed by 35 cycles of 95°C for 15s, the appropriate annealing temperature (see Table 2) for 30s , and 72°C for 1 minute. The final extension was at 72° C for 10 min. and then a final hold at 4°C.

Confirmation of Expected PCR Product by Acrylamide Gel Electrophoresis

The PCR products were run on an 6% acrylamide gel (4.22 mL dH_2O, 1.3 mL 5x SB Buffer, 975 µL 19:1 Acrylamide-Bis mix [Bio-Rad Laboratories, Hercules, CA], 16 µL 10% APS [Amresco, Solon, OH],, and 16 µL TEMED [Amresco, Solon, OH]) to confirm the presence of an amplified product of the correct length. Electrophoresis was conducted in the BIO RAD Mini PROTEAN Tetra Cell (Bio-Rad Laboratories, Hercules, CA) at 80 volts for approximately one hour. Gels were then stained in a ~2 µg/mL ethidium bromide (EtBr) solution (Amresco, Solon, OH) for 3 min., then visualized on a UV transilluminator using the AlphaImager program (Alpha Innotech Corp., San Leandro, CA).

Preparing Samples for CheckIT Gel:

The PCR fragments were selected from gels and prepared for sequencing as described by Hellman et al. (2006). Amplification of each sample was conducted in triplicate, using the same PCR conditions and reagent mix as before in order to increase the yields of DNA products. The PCR product triplicates were concentrated using a Savant DNA 110 DNA SpeedVac (GMI Inc., Ramsey, MN) and then re-suspended in 4 µL of dH_2O to meet the manufacturer's (CheckIT gel) recommendation to use only 4 µL of product to 1 µL of the (5x) loading buffer. Some samples required further dilution by adding 2 µL re-suspended product to 2 µL dH_2O. Samples were run on Elchrom Scientific PCR CheckIT gel with EtBr (Elchrom Scientific U.S.A. Inc., New Hyde Park, NY) using an owl separation system, model B2 (ThermoScientific, Portsmouth, NH). Gels were visualized using a UV light and the Alpha Imager program. A felt-tipped pen was used to mark the plastic backing to identify the bands that would be punched using an Elchrom Band-Pick punch (Elchrom Scientific U.S.A. Inc., New Hyde Park, NY). The punches, representing 2 µL of DNA, were then removed and amplified using the same PCR conditions as described above.

PCR Clean-up and Sequencing Preparation:

The PCR product was cleaned up by adding 5 µL of the product to 2 µL of the ExoSAP-IT (USB Corp., Cleveland, OH). This mixture was then placed in a thermal cycler at 37°C for 15 min., followed by 80°C for 15 min. One µL of the Exo-SAP product was added to 2 µL BigDye Terminator v3.1 ready reaction mix (Applied Biosystems), 2 µL BigDye sequencing buffer, 3.2 µL 1 µM Primer, and 11.8 µL of dH_2O. Separate reaction mixes, each using either the forward or reverse primer, were placed in a PTC 100 Thermalcycler at 96°C for 1 minute, followed by 25 cycles of 96°C for 10s , 50°C for 5s, and 60°C for 4 min., followed by 4° C for 10 min. Ten µL of the BigDye Terminator product were combined with 45 µL SAM solution and 10 µL BigDye XTerminator Solution (Applied Biosystems) in a 96-well sequencing plate and capped. The plate

was shaken on a tabletop vortex for half an hour to thoroughly mix all the solutions, centrifuged on a Beckman model TJ-6 for 5 min. at approximately 2500 rpm (Beckman Coulter Inc., Fullerton, CA) and run on the ABI 3130 (Applied Biosystems).

Data Analysis:

Sequencing data were analyzed using the Sequencher 4.7 DNA sequence editing program (Gene Code, Ann Arbor, MI). Here, the forward and reverse sequences were combined to form contigs, and the sequence data were evaluated and manually edited. Once the sequences were determined, they were exported in FASTA format from the Sequencher program and loaded into the BioEdit program where they were aligned for comparison. In some cases, excessive variation within the repeat regions precluded alignment using BioEdit, requiring the number and motif of repeats to be determined manually. In accordance with the ISFG, amplicon sequences were read in the 5' to 3' direction and for sequences that have been already described in the literature, the repeat sequence motif that was first reported was used as a basis for nomenclature (Bär et al. 1997). For loci whose sequences were not previously reported, the repeat motif was designated based on the first 5' nucleotide that defined the motif.

III. Results

a) Population genetics study and genotype database development (Kanthaswamy et al., submitted).

The number (n), range, and frequencies for each allele in each of the four U.S. geographic regions and the nationwide sample set (the sum of the four regions) are presented in Table 3. The number of alleles per locus ranged from 6 to 29 with a mean of 14. Two-way contingency chi-square $\chi 2$ tests for homogeneity of allele frequencies among all four regional populations showed that frequency distributions differed significantly ($p \leq 0.01$) at only three (i.e. FH2010, FH2309, and PEZ05) of the 18 loci studied. Contingency tables for pairs of geographic locations showed that between 0 and 2 loci showed significant allele frequency differences at the $p \leq 0.01$ level. Therefore, allele frequencies did not differ considerably between regions supporting the pooling of regional data and substructure correction concomitant with the pooled data.

The highest pairwise LD was observed in the mid-west and the lowest LD was observed in the northeast. In the regional samples, 56 and 36 pairs of loci were significantly associated at the 5% and 1% probability levels, respectively. Among the 13 pedigreed and mixed breed dog populations from 3 (in Beagles) to 19 (in mixed breed dogs) pairs of loci were significantly associated at the 5% level of probability while only one pair of loci (in Beagles, Miniature Poodles and Rottweilers) and 5 loci (in mixed breed dogs) were significantly associated at the 1% level of probability. Each locus pair assayed for the pedigreed dogs was statistically significant 5 times each at the 5% and 1% levels, respectively, while the mixed breed dogs were statistically significant 19 and 10 times at the 5% and 1% levels, respectively.

Four pairs of physically linked STRs (FH2107 and FH3377, FH2054 and PEZ05, FH2017 and FH2088, and PEZ16 and vWF.X), which are located on chromosomes *Cfa* 3, *Cfa* 12, *Cfa* 15, and *Cfa* 27, respectively, did not exhibit greater statistically significant association of genotypes at the $p \leq 0.05$ or $p \leq 0.01$ levels of probability than unlinked loci. Therefore, despite their synteny, these markers were treated similarly to the biologically independent loci and were included in all subsequent population genetic analyses.

Fisher's (1932) exact tests across loci within breeds (data not shown) indicate highly significant Hardy Weinberg (HW) disequilibrium within loci in American Pit Bulls, Golden Retrievers, and the mixed breed dogs. Across breeds, the loci FH2017 and FH2107 were observed to be in highly significant HW disequilibria, which suggests non-random associations among alleles at these loci only. The nationwide mean observed and expected heterozygosities

were 71% and 79%, respectively. Observed heterozygosity at each geographic location ranged from 68% to 73% while the estimated heterozygosity (or gene diversity) ranged from 78% to 79%. Across locations, gene diversity estimates were systematically greater than observed numbers of heterozygotes. The data support that while there is a decrease in observed heterozygosity, a high degree of diversity exists within and among breeds and geographic regions.

Breedwise observed and expected heterozygosities mean values ranged from greater than 50% (in German Shepherds and Rottweilers) to 69% and 77%, respectively, in Dachshunds. Estimated heteozygosity exceeded observed heterozygosity in all breeds except German Shepherds where these values were approximately equal. When the estimates from all three Poodle breeds were combined, the observed and expected heterozygosities were 64% and 71%, respectively. The Toy Poodles exhibited significantly greater heterozygosity than the combined estimates. Among the mixed breed dogs the observed and expected heterozygosities were 75% and 79%, respectively.

The Fis estimate, which measures the degree of inbreeding, was somewhat lower among breeds (0.06) than among regions (0.10) while the degree of genetic differentiation (or genetic subdivision) among the breeds (Fst or fixation index) was much higher (0.09) than that among regions (0.002). Fit, which reflects the combined effects of inbreeding and genetic subdivision, was higher among the dog breeds including mixed breed dogs (0.14) than among the four geographic regions (0.11).

The mean genetic differentiation (pairwise Fst) among populations of dogs from each U.S. region ranged from 0.0006 between the mid-west and the south, to 0.0039 between the west and the northeast. The genetic differentiation of the pairwise Fst comparison between breeds ranged from 0.02 (Toy and Standard Poodles) to 0.2788 (Rottweilers and German Shepherds).

Our results show no significant differences between estimates of allele frequency, heterozygosity, and inbreeding coefficients based on all linked and unlinked loci and estimates based on only the most informative unlinked loci (i.e. without loci FH2017, FH2107, vWF.X, and PEZ05, each of which had exhibited lower heterozygosity values compared to the locus to which it was physically linked).

Results of the STRUCTURE analyses of the regional samples and each of the 13 pedigreed and mixed breed populations are illustrated in Figures 3 and 4. These results are concordant with estimates of the allele frequencies, observed and expected heterozygosities, and F-statistics. Individual assignment tests revealed high assignment success for the purebred dogs with all samples being assigned to their correct reference populations. The average within-breed assignment success score ranged from 99.17% (Poodle) to 100% (Dachshund, Golden Retriever, Labrador Retriever, Rottweiler, Shi Tzu and Yorkshire Terrier). In the mixed breed population, the average assignment score was 97.19% with a range of 38.19% to 100%. However, 35 of the 431 mixed breed dogs were assigned with a higher likelihood to one of the purebred populations instead of the mixed breed population. The regional PCA reveals no geographic distribution of variation among the domestic dogs in the U.S., while the breed PCA demonstrates substantial differentiation among these dog breeds with German Shepherds appearing as outliers.

b) Development validation studies (Dayton 2008)

Sensitivity Study

Varying amounts of template DNA from two dogs were used to test the reagent kit's sensitivity. DNA was extracted from whole blood samples from a Golden Retriever and an Akita. Serial dilutions of both samples (10, 5, 2, 1, 0.5, 0.25, 0.125 and 0.0625ng) were amplified in duplicate. All PCR products were diluted 1:10 prior to loading 2uL of sample into 10uL of 0.02% LIZ/formamide solution for detection on the 3130 Genetic Analyzer. This was done to reduce off scale data for the 10ng template amplified samples. For both the Golden

Retriever and Akita samples, all peaks were detected when concentrations of template DNA ranging from 10-0.125ng were used for amplification. One Golden Retriever sample that used a concentration of 0.0625ng template DNA exhibited a true homozygous peak at the FH2017 locus that was below the interpretational threshold of 50 RFU. In the duplicated Golden retriever sample amplification, one of two heterozygous peaks at the FH2004 locus was below the interpretational threshold. All peaks were called (i.e., RFU>50) for the duplicated Akita sample amplifications at every template concentration (including the 0.0625ng samples). Because a 1:10 dilution was made for all PCR products prior to loading on the Genetic Analyzer, on a separate injection plate, 1uL of non-diluted PCR product was pipeted into 10uL of 0.02% LIZ/formamide solution for all template amplification amounts. Similar peak dropout results were obtained when no PCR product dilutions were made. On average, the 1:10 diluted samples had RFUs that were 52% that of the non-diluted samples with a 9.7% standard deviation. When no PCR product dilutions were made, increased pull-up artifacts were seen for the 10 ng template samples. Blue into green pull-up peaks ranged from 2031 to 3774 RFUs. Elevated baseline noise was also present in the green and yellow dye channels for the 10 ng template samples and non-template adenylation (27.4% of the parent peak) was observed for marker FH2001 in the Akita sample. Blue into green pull-up artifacts were also seen for the 5 ng template samples. The off-scale and pull-up peaks indicate that 10 ng of high-quality DNA is too much for the Finnzymes multiplex.

Peak height ratios (PHR) for several markers (FH2328, FH2309, PEZ05, FH2010, vWF.X, FH3313, and FH2088) started to fall below 60% when DNA template amounts of less than or equal to 0.25ng were amplified. Noticeable stochastic fluctuations were apparent for the FH2004 locus, which experienced allelic drop out in one of the 0.0625ng amplifications. The other amplifications with 0.0625ng template DNA exhibited a PHR range between 41.5-63.2% (n=3).

To verify that the template amounts prepared in serial dilutions were the amounts used in the experiment, the diluted samples were re-quantified using PicoGreen®. The minimum Akita sample (0.0625ng) was 0.01 fluorescent units higher than the Golden Retriever sample when quantified using K562 DNA standards. The 0.125ng samples were quantified at the same fluorescent values meaning the serial dilution precision for the Golden Retriever and Akita samples were better for the 0.125 ng DNA template compared to the 0.0625 ng DNA template dilutions. At 0.125 ng template, the Finnzymes multiplex was capable of detecting all profile peaks at >50 RFU (all but one peak at the FH2017 locus [63 RFU] are above 100 RFU) when no PCR product dilutions were made prior to adding sample to the LIZ/Formamide solution. This level of sensitivity is comparable to that of previous STR multiplexes developed. For instance, the 'Meowplex' produces full STR profiles from 125pg of template DNA (Menotti-Raymond et al. 2005), while the three canine STR multiplexes (MP1, MP2 and MP3) developed by Eichmann et al. (2005) are capable of producing full profiles from 100pg for MP1 and MP3, and 250pg for MP2. A similar input DNA template range of 0.5-1.25 ng is also seen for the Identifiler® kit (Collins et al. 2004). In addition, while the largest number of markers co-amplified in the afore-mentioned studies is 15, a unique feature of the Finnzymes multiplex is that it includes a total of 19 markers.

Reproducibility

Five canine International Society of Animal Genetics (ISAG) standards were obtained as pre-quantified genomic DNA samples (http://www.isag.org.uk/ISAG/all/ISAG2006_CompanionAnimals.pdf). The samples were amplified and run at both the Molecular Anthropology Laboratory (MAL) at the University of California, Davis and at the Jan Bashinski DNA Laboratory, California Department of Justice, Richmond, CA. At the MAL, an Eppendorf Mastercycler epigradient thermocycler (Eppendorf, Hamburg, Germany) was used to amplify the ISAG samples, and an ABI 3130 Genetic Analyzer containing POP-7™ (Applied Biosystems, Foster City, CA) was used to resolve the amplified

fragments. At the Jan Bashinski DNA Laboratory, a GeneAmp® 9700 PCR thermocycler (Applied Biosystems, Foster City, CA) was used to amplify the samples and an ABI 3130 Genetic Analyzer containing POP-4™ (Applied Biosystems, Foster City, CA) was used to resolve the amplified fragments. All samples were amplified in duplicate. All sizes were calculated using a global southern method with the 75, 100, 139, 150, 160, 200, 300, 350, 400, and 450bp LIZ peaks.

The samples resolved with POP-7 ran through the 3130 capillary much faster (75bp LIZ peak resolved at ~1500 data points) as compared to the samples resolved with POP-4 (75bp LIZ peak resolved at ~2600 data points). Variations from 2-5bp occurred between ISAG samples run at the different laboratories. This may be partly due to the difference of mobility between allele peaks and GS-500 peaks based on ambient room temperatures and because different sieving mediums (POP-4 versus POP-7) were used. At lower ambient temperatures, STR alleles tend to migrate slower than the GS-500 fragments and at higher temperatures the alleles migrate faster than the GS-500 fragments, causing differences in the sizing of those STR fragments (Klein et al. 2003). The canine control DNA included in the kit can be used to calibrate the allele sizes observed. However, due to the presence of only 1-2 alleles per locus, calibration may be inaccurate, particularly if a locus spans a large size range. An allelic ladder would mitigate the effects of migration variations between laboratories. Accurate concordance data (binned within 1 bp) is difficult until the allelic ladder is fully developed and utilized in subsequent runs. Currently, a canine allelic ladder is being developed for the Finnzymes multiplex, but without the use of a ladder, consistent genotyping of PCR product peaks on a capillary electrophoretic platform may be difficult.

Intracolor and Intercolor balance

The intracolor percentages were calculated for 61 samples. Products in the red channel consistently produced the best intracolor balance among loci at 45.18% with a standard deviation of 12.85 (n = 61). All other dye colors had intracolor balances ranging from 32.77±10.5% (green), 33.63±14.18% (blue) and 38.69±12.92% (yellow). For comparison, the amplification of non-degraded and uninhibited samples of human DNA using the Identifiler® kit was reported to attain an average intracolor balance of 50% or higher (Collins et al. 2004). Intercolor balance for 61 samples was estimated to be 21.97±7.67% which is within the lower range of intercolor balance for the Identifiler® kit (20-40%). The reduced color balances of the canine STR kit may be attributed to particular marker primer pairs.

All loci that exhibit the lowest or the highest normalized RFU values are possible candidates for further primer concentration studies to improve intracolor and intercolor balance. In the blue channel, locus FH2017 demonstrated the lowest RFU value for 36 out of 61 samples, followed by locus FH2309 which had the lowest RFU value for 16 samples. The low RFU values for locus FH2017 might be partially explained by a suspected primer binding site mutation. The Zinc Finger locus usually exhibits the highest RFU value within the blue channel, but locus FH2017 also had the highest RFU value in 11 samples (presumably there is no primer binding site variant in these samples). The locus with the lowest RFU value in the green channel is PEZ05 differing 1000-3000 RFU from the highest normalized RFU in the green channel. In well balanced primer mixtures, primer pairs that produce the smallest PCR product lengths would be expected to produce the highest RFU signal because it is easier for the polymerase to amplify smaller products (Leibelt et al. 2003). Because locus PEZ05 (the smallest PCR product in the green channel) exhibits the lowest RFU value, the current primer pairs for PEZ05 are not performing optimally with the other primers in the same dye channel. Interestingly, three different markers (FH2001, FH2004 and FH2361) competed for the highest RFU's in the green dye color. Thus, if the signal from locus PEZ05 can be increased in relation to the other markers labeled in green the intracolor balance can be improved. For the yellow channel locus FH2107 tends to produce the lowest normalized signal while locus FH2054 produces the highest signal,

the difference being 1000-1700 RFU. The red channel exhibits the best averaged intracolor balance of 45.18%. Loci FH2088 and FH3313 tend to produce the lowest RFU signal, 1000-2000 RFU lower than normalized peak produced by locus FH2010 (with the highest RFU peaks).

Amplification-Annealing Temperature

When multiple primer sets are co-amplified in a single PCR reaction, variation in locus product yield will vary somewhat under one annealing temperature. It is important to validate how prone a reagent kit is for variation in annealing temperature, because thermal cyclers may exhibit some variation between and within their thermal blocks. Therefore, different annealing temperatures were evaluated in order to investigate the amplification success of the entire suite of primers in the Finnzymes multiplex as a function of annealing temperature. For this study, three DNA samples (2 bloods and 1 buccal extracted sample) were amplified in duplicate. One μL of PCR product was added to 10 μL of 0.02% LIZ/Formamide solution for electrokinectic injection. The best overall intracolor balance occurred at 60ºC, with average color balances being 42.41% for blue, 52.99% for green, 53.31% for yellow, and 54.98% for red. A 60ºC annealing temperature also provided the best intercolor balance at 29.0% compared to 27.0% for 58.5ºC and 22.0% for 57.7ºC. Between 56 ºC and 60 ºC, all allele peaks were called; the PEZ05 locus dropped out in one amplification at 61.9 ºC and had RFU values below 100 for the other five amplifications (compared to RFUs of 500-4000 for all other loci). The manufacturer of the kit recommends an annealing temperature of 60 ºC, which is congruent with the results herein.

Amplification-Number of cycles

To determine the optimal number of cycles for the canine STR kit, PCRs (1ng template in a 20uL reaction) were run for 26, 28, 30, 32, and 34 cycles on a GeneAmp 9700 thermocycler (Applied Biosystems, Foster City, CA). DNA from two blood extracted samples and one buccal extracted sample were amplified in duplicate. The manufacturer (Finnzymes Oy, Espoo, Finland) recommended 30 cycles for amplifying canine DNA. As expected, the overall signal intensity for most peaks increased with cycle number.

Four of the six amplifications exhibited allelic dropout at locus FH2017 after only 26 PCR cycles, but all other cycle numbers provided peaks above 50 RFU. The best overall intracolor and intercolor balances were seen at 28 and 30 cycles. Between 26 and 32 cycles, signal intensity increased with increasing cycle number, but at 34 cycles certain loci had considerably higher signal than others. Peaks for the Zinc Finger, FH2017, FH2088, FH2010, and vWF.X loci had the highest signals (5000-8000 RFU) while most other peaks had signals below 2000 RFU. As expected, no shouldering from non-template nucleotide addition was seen at any of the cycle numbers tested because unlike Taq polymerase, Phusion Polymerase has 3' to 5' exonuclease activity that eliminates the non-templated nucleotide additions.

Peak Height Ratios

Sixty one samples were amplified to calculate peak height ratios (PHRs) for heterozygous allele pairs. Only samples that amplified peaks exceeding 500 RFUs were used in the calculations. PHRs (excluding FH2017 and Zinc Finger) averaged between 78.57 and 96.14%. The FH2017 locus exhibited a standard deviation in PHR over three times that of the locus with the next highest standard deviation. Also, the low heterozygosity (0.49 for n=61) for locus FH2017 in relation to other markers in the panel suggests a primer binding site mutation may be causing allelic imbalance. Another indication of a primer binding site mutation at locus FH2017 is the observed increase in PHR as the annealing temperature is decreased (Butler 2005). One sample from a Labrador retriever used in the annealing temperature study also exhibited this phenomenon. Eleven other samples produced low peak height ratios ranging from 22.41% to 8.32%. One solution to address low peak signal caused by primer binding site mutations is to design a degenerate primer for FH2017 to add to the PCR reaction (Leibelt et al. 2003; Butler

2005). The primer binding regions for animals demonstrating FH2017 binding mutations should be sequenced to determine the location and base composition of the mutation. Because both intracolor and intercolor balance are poor at 56ºC, reducing the stringency of the PCR by reducing the annealing temperature is not a feasible option.

In addition, the Zinc Finger locus exhibited a lower peak height ratio average (51.20%) with the male specific peak (~164 bp) amplifying 1000-2000 RFU shorter than the X specific peak (~159 bp). All samples that contained sex information about the donor animal were typed correctly by the Zinc Finger locus (n=67).

Stutter Percentages

The stutter percentages were calculated by dividing the RFU value of the stutter peak (n-4) by the RFU value of the parent peak (n) when the parent peak exceeded 1000 RFUs. Peaks with a heterozygous allele in the n-4 stutter position were not used to determine the mean stutter percentages for the locus. Locus PEZ17 contains both n-2 and n-4 stutter peaks. The average n-2 stutter percentage for PEZ17 is 12.10% with a standard deviation of 3.59% (n=75) when the parent peak is taller than 1000 RFU. Locus FH2309 sometimes exhibits n+4 stutter peaks (n=3). The height of the FH2309 parent peak was typically above 2000 RFU when n+4 stutter was seen with one exception of a parent peak at 900 RFU that contained an n+4 stutter artifact.

Artifacts

Several dye blobs were seen in the reagent blanks and in samples containing peaks with low RFU values. Dye blobs are seen at ~123 bp in the blue dye channel, at ~117 bp in the green channel, at ~96 bp in the yellow channel, and at ~112 bp and ~118 bp in the red channel. The dye blob in the yellow channel sometimes interfered with the allele calls in marker PEZ21 if the peak height for the allele was at or below 100 RFU.

Mixture Study

A binning program called FLEXIBIN (Amos et al. 2007) was used to determine allele calls for the animals of the mixture study (Kanthaswamy et al. submitted). The ‘Repeats’ column is not an accurate reflection of the number of tetranucleotide repeats present in the actual PCR product but the values are operationally defined.

Two genomic canine samples were mixed in the following proportions so that the total DNA input for each reaction was 1.0 ng: 20:1, 10:1, 5:1, 2:1, 1:1, 1:2, 1:5, 1:10 and 1:20. For this study, loci that exhibited non-overlapping alleles (as determined by the unmixed controls) were selected. The minor component robustly amplifies (all peaks above 50 RFU) at a 1:5 mixture ratio. This may be partly due to lower intracolor balances across all dye channels. The Identifiler® kit typically has an intracolor balance of 50% for all color channels when samples are not degraded or inhibited, which may explain the typing kit’s ability to robustly amplify the minor component in mixture ratios of 1:10 (Collins et al. 2004). Better detection of minor component DNA in canine mixtures might be attained by improving intracolor balance (see ‘Intracolor and Intercolor balance’ section).

Overlapping Marker size ranges

Size range overlap among loci was detected in the electropherograms of 67 different California dog samples. A marker range overlap of 14 base pairs occurred between FH3377 and FH2107 in the Akita sample from the sensitivity study. The overlap was only detected because the neighboring marker range for FH2107 contained two heterozygous peaks that are of significantly less intensity than the homozygous FH3377 peak. Of 67 animals, the Akita sample was the only example of an allele size in the FH3377/FH2107 overlapping region. The analyst should be cautious when making an interpretation of an allele in the overlapping range for FH3377 and FH2107 because both markers contain an allele ~ 295.6 bp. All other alleles that

type in the overlap region should be more easily distinguished with the binning properties of the different length repeat motifs since FH3377 is a pentanucleotide repeat and FH2107 is a tetranucleotide repeat. Another overlapping allele range occur between markers FH2004/FH2361 that share a 3-4 base pair overlap, but perfect repeat allele bins in both marker ranges are offset by one base pair from each other.

Microvariants Seen in FH3313 and FH2361

Three heterozygous samples typed in this study exhibited a microvariant allele for locus FH2361. One animal exhibited heterozygous alleles one base pair apart, and two other animals exhibited heterozygous alleles two base pairs apart. The peaks that are a single base pair apart are resolved clearly. Five animals contained microvariant alleles in the FH3313 locus. Three animals contained well-resolved FH3313 microvariants which appeared to be two base pairs apart, but two other canine samples had broad peaks. The samples containing broad peaks were confirmed by a second amplification and two separate injections. The morphology of these peaks might be caused by STR alleles of the same length on either autosomal chromosome containing slightly different sequence variations from each other, resulting in different migration properties during electrophoresis (Rosenblum 1997; Butler 2005). Microvariants were also observed at loci FH2107, FH2309, and FH3377 when an independent larger sample set of 667 dogs was analyzed (Tom et al. submitted; Kanthaswamy et al. submitted).

Species Specificity

The Finnzymes multiplex amplified reproducible STR profiles in 5 wolf (*Canis lupus*) samples as would be expected given the evolutionary history of the domestic dog and wolf (Brewer et al. 2002). The peaks obtained in wolves were concordant with the allele spectrum of the domestic dog for each locus. Indeed, STR loci are well-known to cross-amplify even among more genetically differentiated species than the dog and wolf (Primmer et al. 1997). For example, the Identifiler® kit developed for human profiling generates partial profiles for other primate species (chimpanzee, orangutan, macaque, and gorilla, Collins et al. 2004), and the "Meowplex" generates PCR products for ocelot, and puma samples (Menotti-Raymond et al. 2005).

Therefore, canine species specificity of the 19-plex was assessed by amplifying 5ng non-canine DNA samples in duplicate. Several peaks were observed when DNA samples from chicken, mouse, rat, horse, cow, pig, cat, fish (Catfish) and monkey (African green monkey) were typed.

Only low level peaks below 100 RFU were present for both amplifications with rat and fish DNA. Peaks above 50 RFU were seen in one of two amplifications of pig DNA. Most peaks fell outside of the canine allele ranges or were incongruent with the ladder positions while a few non-canine specific peak sizes were close to canine allele positions. To better assess whether these peaks truly fall into a canine allele bin, the various non-canine species (fish, pig, rat, cat, mouse, chicken and monkey) should be run again with an allelic ladder, once it is developed. Subsequent BLAT and BLAST searches of the 19 sets of primers revealed that some primers have 95-100% identity with the genomic information for certain species; however, no pair of primers (both forward and reverse) for a given locus demonstrated such high similarity. Primers FH2361R, vWF.XF, and FH2017F have the most similarity with a number of non-canine species. The PEZ17R primer has 100% identity with 100 locations in the cat genome. But, the primer binding sites discovered with the BLAT and BLAST searches for the 19-plex primers do not assess the possibility for non-target binding of the 3' end of the primers. For instance, primer binding to chicken, cow, and pig samples was not confirmed by searching for similar DNA sequences between the Finnzymes multiplex primer panel and the genome information available on the BLAST and BLAT websites. Similarly, several canine primers shared 95-100% identity with multiple locations in the human genome, but amplification peaks were not present in

duplicate 19-plex reactions of human DNA (data not shown). A simple BLAST or BLAT search may be insufficient to assess cross-reactivity of a primer to a particular genome. Also, none of the peaks exhibited in any of the non-canine species appeared to reveal the morphology of STR products (all peaks above 500 RFU lack stutter peaks), which could be helpful in evaluating whether non-canine species amplification has occurred in a reaction. Of course, species cross-reactivity does not necessarily make a kit unreliable. It is important to perform such studies to better understand the limitations of an assay. No full 19 locus profile was observed with any of the above mentioned species.

Case-Type Samples

Samples from a recent dog attack were amplified in duplicate using the Finnzymes multiplex. Sample numbers 1 through 13 were single shed hairs found on a sweatshirt and samples 14 through 17 were saliva stain samples found on the victim's sweatshirt. Although not recommended for routine casework, the detection threshold was lowered to 25 RFU for all colors to collect as much data as possible. Using the California Department of Justice casework peak threshold standards, only peaks called above 50 RFU were considered 'alleles', and single peaks at a locus above 200 RFU were considered a homozygous genotype. A ten second injection time at 3 kV was used while running the sample on the ABI 3130 CE. Table 4 presents a combined summary of the peaks called from duplicate PCR reactions; all alleles are described as 'raw' lengths in base pairs. Many of the case work samples had peaks below 50 RFU, and many markers only had a single peak call possibly indicating peak drop-out due to low input DNA amounts amplified in the PCR reactions.

Samples 8 and 13 had three peaks in the marker range for PEZ17, some of which were below 50 RFU. Some of the peaks below 50 RFU may be attributed to baseline noise or allelic drop-out affected by stochastic fluctuations from typing low amounts of genomic DNA (Butler 2005). Sample 8, a single shed hair extracted using the Chelex method, provided the best profile overall with the most peaks called above 50 RFU. Additional technical modifications will be required to amplify samples below 0.125 ng such as increasing the cycle number and investigating alternative extraction protocols for telogen hair samples (Braunner et al. 2001; Lu et a. 2003), but these actions carry all the limitations experienced with low copy DNA analyses (Budowle et al. 2001). Based on the positive control sample that was processed on the same injection plate as the non-probative samples, allele calls for sample 8 were obtained by binning (for peaks above 50 RFU (Table 5)).

The Fst and the allele frequencies were determined from a 667 animal population study from canines across the United States (Kanthaswamy et al., submitted), the RMP for the positive control animal profile was calculated at 3.47×10^{-23}, and a LR at 2.89×10^{22}. The RMP for the sample 8 profile was calculated at 1.00×10^{-17}. Two saliva stain samples (14 and 17) provided partial profiles containing peaks from 10 and 12 loci respectively. Using the Fst value of 0.09, the RMP for the sample 14 profile is 3.18×10^{-3}, and sample 17 has a RMP of 1.04×10^{-3}.

The sex typing marker (Zinc Finger) had varying success when typing telogen hair samples. Of the ten hair samples that produced peaks, three samples typed X and Y associated peaks; however, only two of these samples had a Y peak 50 RFU. In addition, only four hair samples typed the X peak above 50 RFU, and three samples had X peaks between 25 and 50 RFU. As seen with the PHR study, the ~159 bp X peak typed at much higher RFUs (typically two times higher) than the ~164 Y peak in male canines, and, therefore, the ZF-X primers seemed to amplify more robustly at input DNA levels below 125 pg. The two saliva stain samples that generated profiles amplified both the X and Y peaks above 50 RFU.

Positive Control Animal

Interestingly, the positive control F-863 produces a profile that contains three peaks for marker FH2328. Originally, it was suspected that F-863 contained rare alleles for PEZ05 and

FH2001 because this would account for all six peaks seen amongst neighboring loci PEZ05, FH2001 and FH2328. To test this theory, F-863 was amplified with single plexes of PEZ05, FH2001 and FH2328. In triplicate reactions, 4.5 µL each of the forward and reverse primers (500 nm total) was combined with 9 µL of the Finnzymes Reaction Mixture and 2 µL for each single plex. The results revealed that the positive control animal produced a single peak of ~100.4 bp at the PEZ05 locus, two peaks of ~128.01 bp and ~144.8 bp at the FH2001 locus, and two peaks of ~169.9 bp and ~205.5 bp at the FH2328 locus. In two of the three amplifications, FH2328 produced what appeared to be a third peak of ~201.6 bp. All of the amplifications of the positive control animal using the Finnzymes multiplex throughout the validation (n=13) contain the third peak (~201.6 bp), averaging a height of 53.0% that of the ~205.5 bp peak. Using Picogreen®, the positive control animal was quantified at 0.1 ng/µL, meaning that all of the positive control animal amplifications were performed using only 0.2 ng of DNA in a 20 µL reaction. This may explain why the third peak was only seen in two of the three single plex amplifications using the FH2328 primers. This third FH2328 peak could possibly be an artifact peak generated from the 19-plex or it could be a tri-allelic pattern inherent to the cell line used for the positive control. The Zinc Finger sex typing marker successfully typed the positive control animal as female in all ten amplifications. Typically, the ~159 bp X peak was the tallest peak in the blue color channel for the positive control.

The proposed multiplex is comprised of 18 STR markers and a zinc-finger sexing marker. Five of the 18 STRs included in the proposed panel have been described previously, and the core repeat units of their common alleles have been published in detail (Eichmann et al. 2005; Hellmann et al. 2006). Of these five, FH2010, FH2054, FH2328, and PEZ02 were tetrameric while the VWF.X marker was a hexameric repeat unit. Although the tetrameric repeat units of nine of the remaining loci have been published, detailed characterization and nomenclature has not yet been developed. These loci include Pez05, Pez16, Pez17, Pez21, FH2001, FH2004, FH2088, FH2017, and FH2107 (Shibuya et al. 1994; Francisco et al. 1996; Neff et al. 1999; Halverson et al. 1999; Mellersh et al. 2000; Breen et al. 2001; Guyon et al. 2003). The repeat units and nomenclature of these loci as well as those of FH2309, FH2361, FH3313, and a pentameric locus, FH3377, are described in detail in the present research.

Repeat unit structures were ascertained by aligning sequences and determining differences in the number of repeats within the core repeat region. In a few cases, complete sequences could not be obtained due to suboptimal priming sites, and, thus, similarity was confirmed by aligning the partial sequences obtained in this project with those provided by W. Parsons and H. Parker and sequences published by Eichmann et al. (2005) and Hellmann et al. (2006). Unpublished sequences were confirmed by aligning sequences obtained here with those of the published canine genome (http://genome.ucsc.edu/).

The primary repeat motif (i.e. the most common motif within a sequence) type, length, and sequence for each STR locus are listed according to ISFG recommendations (Bär et al. 1997). For the compound and complex loci, the first observable and primary repeat motifs are not necessarily the same. The observed repeat lengths of most loci were concordant with the estimated repeat unit lengths generated by the program. FH3377 exhibited a unit length of 5, and VWF.X exhibited a unit length of 6 while the other loci exhibited a repeat unit length of 4.

For loci FH2309 and FH2361, there were several instances where the actual allele sizes (*Allele*), which are based on the sequences, did not correspond with the estimated allele bin length (*Length*) based on genotypes from a capillary electrophoresis platform. Eichmann et al. (2005) also observed FH2079 amplicons which showed no nucleotide sequence differences but migrated at different rates. Conversely, amplicons exhibiting the same electrophoretic mobility rate and sharing the same sequence size but containing different repeat patterns have been reported in dogs and humans (Hellmann et al. 2006). Although mutations in regions flanking the core repeat structure undetectable by fragment sizing (Dayton 2008) may account for these discrepancies, additional sequencing would help to clarify this issue. These observations suggest

that absolute reliance on amplicon size for each of the possible alleles may be misleading. Loci FH2309 and FH2361 also contained a deletion or an addition of a tetrameric sequence in their flanking regions outside of the repeat region causing changes to the allele size, and, therefore, allele number as noted by parentheses in the tables below. Of these six loci, only FH2017 did not exhibit any alleles containing microvariants (or incomplete or partial repeat motifs). Dayton (2008) has reported heterozygote dogs with alleles one base pair apart at locus FH2361, and microvariants in FH3313 that appear to be two base pairs apart. They also observed mutations in the priming sites of locus FH2017 that may have caused allelic dropouts and therefore lower estimates of heterozygosity at this locus compared to other loci used in a population genetic assessment of 667 dogs (Kanthaswamy et al. submitted).

The Eichmann et al. method for nomenclature development included the entire repeat region in the calculations of the allele number (2005). This method was used when assigning allele numbers at each locus allowing for the inclusion of novel alleles into the nomenclature system. The sequences obtained in this study were analyzed using the model described in Barber et al. (1996). Working from the outside in, the primer regions were analyzed first, followed by the flanking regions, and the core repeat region in the center.

c) Nomenclature development (Tom et al., submitted)

Simple Repeats:

Nine of the 18 STR loci were characterized by simple repeat structures. Therefore, a designated allele number directly represents the number of repeats units within each allele's sequence. The loci Pez02, FH2010, FH2054, FH2328, and VWF.x have all been characterized in detail by Eichmann et al. (2005) and Hellmann et al. (2006). Since primers used in this study annealed at different binding sites from those in previous studies, it was important to verify that observed repeat units were the same as those described for the five structurally characterized loci.

The complete sequences of Pez05, Pez17, and FH 2001 alleles are available in GenBank under the following accession numbers: FJ031004, FJ031006 and FJ030995. Each of these loci also contained simple repeat structures, and, as in the previous five loci, an allele number assigned to each allele represented the number of complete repeat units found within the allele's sequence. According to our analysis, Pez21 also contained a simple repeat motif of AAAT. However, it was difficult to obtain complete sequences for this locus due to the location of the primer sites. Both the forward and reverse primers are located immediately before and after the core repeat region making sequencing of this locus very difficult. Sequencing regions immediately downstream of the Pez21 primers usually resulted in indistinct peaks for the first few bases before clear peaks were observed. While this made the alignment of complete forward and reverse sequences using the Pez21 primer set difficult, the partial allele sequences of this locus that we managed to obtain confirmed the simple repeat structure of the locus. This was verified using complete sequences provided by our collaborators. A partial sequence of one of the Pez21 alleles is available on GenBank under the accession number FJ042512.

The repeat motifs of loci Pez05 and Pez17 observed here were different than those reported by Halverson et al. (1999). In accordance with International Society of Forensic Genetics (ISFG) recommendations, the first observable repeat motif of Pez05 core repeat region that proceeds in a 5' to 3' direction was a TTTA unit and not an AAAG repeat unit as reported in Halverson et al.'s report (1999). For Pez17, the first repeat unit observed was GAAA instead of the AAAG repeat reported by Halverson et al. (1999). Therefore, in accordance with ISFG recommendations, the GAAA motif for the locus' nomenclature was used as the repeat motif for this locus. The NIST webpage (http://www.cstl.nist.gov/strbase/dogSTRs.htm) also supports the TTTA and GAAA motifs of the Pez05 and Pez17, respectively.

FH2001 was a complicated locus as it contained four different tetrameric repeat units (GATA, GGTA, CAAT, and AGAT) proximal to the forward primer with GATA being the first and therefore the primary repeat motif for nomenclature purposes. Though this could be viewed as a compound repeat, only a stretch of GATA units varies among allelic sequences, making it a simple repeat in terms of nomenclature development.

Compound Repeats:

Pez16, FH2017, FH2088, and FH2309 contained compound repeat structures, i.e. exhibited different tetrameric repeat unit motifs. The complete sequences of representative Pez16, FH2017, FH2088 and FH2309 alleles are available in GenBank under the accession numbers FJ031005, FJ0310997, FJ030998 and FJ031000, respectively.

Pez16 exhibited a compound repeat structure that consisted of two different repeat units, GAAA and GGAA instead of an AAAG repeat unit reported by Halverson et al. (1999). In accordance with ISFG, the GAAA unit was defined as the primary motif as it was the first observable repeat unit, a conclusion also supported by the NIST Dog STR webpage (http://www.cstl.nist.gov/strbase/dogSTRs.htm). Among the Pez16 alleles that were sequenced, the two smaller alleles exhibited differences in the GGAA unit while the GAAA unit differed among the three larger alleles. This was interesting because despite the motifs that differed among the smaller and larger alleles, all these alleles were sequentially ordered 16, 17, 18, 19, and 20.

FH2017 exhibited a compound repeat motif with three tetrameric sequences that followed one after another: AGGT, AGAT and GATA with GATA being the primary repeat motif seen in the repeat sequence. Between 10 and 14 GATA repeat units were observed among the alleles sequenced followed by 2 AGAT units, which were followed by either 5 or 6 AGGT repeats. The nomenclature for this locus is based on the number of GATA, AGAT and AGGT repeat units found in the entire repeat sequence. Though the AGAT repeat unit did not change in number, it was located between two repeat units that varied in number. Thus, following the rules established for our nomenclature, it was included in the allele number.

FH2088 is another compound repeat that contained two different repeat units, TTTA and TTCA. As in the case of Pez21, this locus was not easy to sequence due to the location of the primer sites near the core repeat region. Using different primer sets that annealed further from the core region, complete sequences were obtained.

FH2309 exhibited compound structure with primary repeat units of GAAA and GGAA. Units of GGGA sometimes started the repeat stretch, followed by GGAA and GAAA, but the GGGA occurred in far fewer numbers than the other two repeat units. Most of the larger alleles contained a polymorphic stretch that did not follow the tetrameric structure. In these larger alleles, following the GAAA's was a GAAAAA unit, followed by another stretch of GAAA. The interpretation of such alleles is problematic as it could be read as a two base insertion or as having this GAAAAA repeat. This GAAAAA unit was usually seen in the longer sequences. The smaller alleles that did not contain the GAAAAA unit had integers for allele numbers, while the larger alleles that had the GAAAAA unit had allele numbers x.2. In addition, a CTGG sequence occurred in a flanking region located 72 bases before the first GGGA stretch. In some sequences, there was a duplication of this CTGG sequence. This polymorphism could cause problems when calling alleles using a ladder as it would not be possible to tell the difference between a GAAA or a CTGG without knowing the sequence. The allele number takes into account the presence of the CTGG tetrameric unit.

Complex Repeats:

FH2004, FH2107, FH2361, FH3313, and FH3377 exhibited a complex repeat structure. The allele sequences for each of these loci contained repeat units that differed in both sequence and length. Therefore, allele numbers were based on the number of full repeat motifs and,

separated by a decimal, the number of nucleotides contained in the partial or incomplete repeat motifs. The sequences for representative alleles of FH2004, FH2107, FH2361, FH3313, and FH3377 are available in GenBank under the accession numbers FJ030996, FJ030999, FJ031001, FJ0131002, and FJ031003, respectively.

According to ISFG recommendations, the primary repeat unit of FH2004 is AAAG, and this differs from the primary repeat unit GAAA reported by Francisco et al. (1996). Sequencing revealed a stretch of A's before the first G in the repeat region. Therefore, the first observable repeat unit would be that of the AAAG instead of GAAA. Interestingly, the NIST webpage describes TTCT as the primary repeat unit for this locus. The smaller alleles observed in our study contained only this AAAG tetrameric repeat unit while in the larger alleles contained the AAAG unit as well as the AAAAAG and GAAG units.

FH2107 was a difficult locus to interpret as it had a polymorphism that complicated the allele calling process. The primary repeat unit at this locus is a tetrameric GAAA repeat. The repeat sequence that was observed started with a stretch of GAAA units. There are three stretches of the GAAA units followed by two constant A bases and then a 12 base stretch that is seen in some sequences and not others. This stretch of 12 bases does not interfere with the nomenclature as it results in a whole number: the twelve bases divides three times into the four base repeat unit.

It was relatively easy to detect the GAAA repeat motif of FH2361. This locus, like the others, had a few problematic regions. One sample exhibited a deletion of a TAGA 48 bases after the last GAAA repeat, which could lead to confusion were it interpreted as a deletion of a GAAA unit. Thus, the allele number that takes into account the addition of this TAGA tetrameric unit deletion is noted in parentheses. Another unique sequence exhibited a GAAAAA prior to the last GAAA unit. Unlike previous loci, the GAAAAA did not occur in a longer sequence and could complicate allele calling. The smaller alleles are all integers while the allele containing the GAAAAA has an off-size allele number *x*.2.

FH3313 is another compound repeat with the primary repeat unit being GAAA. The GAAA repeat unit was seen throughout the repeat structure but was separated by stretches of varying G and A combinations. Two alleles, number 44.1 and 49.1, differed from the consensus by not containing the usual GA-GAAAA-GAAA(n)-GA-GAAA sequence following the third stretch of GAAA's. The third stretch of GAAA's of both alleles were significantly larger than the other sequences being either 16 or 17 units in length compared to the others that were either 5 or 6 repeat units. Allele 44.1 was unique as it was the only allele that had 6 repeat units for the first stretch of GAAA repeats. Also, allele 44.1 was the only allele that contained an α sequence of GAAAA rather than the β sequence of GGGAA, which was seen in all other samples. The α or β sequence appeared after the second variable stretch of GAAA units. Alleles 57.2 and 62.1 also exhibited repeat regions that varied from the others. $GAAA_{(o)}$ was also larger than normal for these two alleles, which were the two largest alleles sequenced. It appears that the composition of the repeat region changes as the alleles increase in size. The two smallest and two largest sequenced alleles exhibit repeat sequences that differ from all other sequenced alleles. The resulting nomenclature did not contain integers but fractional repeat units expressed as *x*.1 and *x*.2.

FH3377 was unique in having a primary pentameric repeat unit of GAAAA. The second most common repeat seen in this sequence was the tetrameric GAAA motif. In some samples, simple tetrameric repeats GAAA(3) are followed by the pentameric repeat GAAAA. Other samples were more complicated since they contained GAAA followed by the GAAAA and also GAAAA-GAAAAAAA-GAAAA(2)-GAAA-GAAAA(2)-GAA-GAAAA(n). Only the larger alleles exhibited this larger, more variable repeat sequence. Treating the shorter, simpler sequences and the longer, more variable sequences separately illustrated the fact that the only repeat unit that changed in number was the pentameric GAAAA repeat. One of the samples containing this large, variable repeat sequence lacked a pentameric GAAAA following the

GAAAAAAA. Another sample containing the shorter repeat sequence contained an additional A before the first GAAA. Since this locus is complex with both tetrameric and hexameric repeat units, the allele numbers are not integers. The allele numbers are based on the pentameric repeat unit, making the allele numbers *x*.2 for the shorter sequences and *x*.3 for the longer sequences.

Sexing Marker

The sex-linked Zinc Finger X (ZFx) and Zinc Finger Y (ZFy) alleles were also sequenced and analyzed. As would be expected, males exhibited both a ZFx and a ZFy band, while females exhibited two ZFx bands. Thus, males are heterozygous at this marker and females are homozygous.

Table 1 Individual Locus STR Information. Chromosomal locations were verified using the UCSC Genome Browser (http://genome.ucsc.edu/). N/A: Information not available Italicized fonts indicate a different repeat motif than that previously reported for a particular locus.

Locus	Reference	Repeat Type	Observed Primary Repeat Motif	Estimated Repeat Length	Effective Repeat Range	Chromosome (Map Coordinates)
FH2001	Francisco et al. (1996)	Tetra	GATA	N/A	N/A	N/A
	Current Research	Tetra	GATA	4.145	118.77-159.97	23 (50961325-50961475)
FH2004	Francisco et al. (1996)	Tetra	GAAA	N/A	N/A	N/A
	Current Research	*Tetra*	*AAAG*	*4.197*	*232.82-325.22*	*11 (32161381-32161621)*
FH2010	Francisco et al. (1996)	Tetra	ATGA	N/A	N/A	N/A
	Current Research	Tetra	ATGA	4.181	221.66-242.66	24 (5196383-5196605)
FH2017	Francisco et al. (1996)	Tetra	$GGTA_{(m)}GATA_{(n)}$	N/A	N/A	N/A
	Current Research	*Tetra*	$AGGT_{(m)}AGAT_{(n)}GATA_{(o)}$	*3.825*	*256.69-275.69*	*15 (37914470-37914741)*
FH2054	Francisco et al. (1996)	Tetra	GATA	N/A	N/A	N/A
	Current Research	Tetra	GATA	4.147	139.09-176.53	12 (37914504-37914739)
FH2088	Francisco et al. (1996)	Tetra	$TTTA_{(m)}TTCA_{(n)}$	N/A	N/A	N/A

	Current Research	Tetra	$TTTA_{(m)}TTCA_{(n)}$	3.971	94.56-138.12	15 (53905651-53905779)
FH2107	Francisco et al. (1996)	Tetra	GAAA	N/A	N/A	N/A
	Current Research	Tetra	GAAA	3.711	291.72-425.64	3 (83830247-83830574)
FH2309	Ostrander et al. (1993)	Tetra	Motif not defined	N/A	N/A	N/A
	Current Research	Tetra	GAAA	3.847	339.66-427.98	1 (85772974-85773377)
FH2328	Hellmann et al. (2006)	Tetra	GAAA	N/A	N/A	N/A
	Current Research	Tetra	GAAA	3.855	171-213.24	33 (19158127-19158477)
FH2361	Mellersh et al. (2000)	Tetra	Motif not defined	N/A	N/A	N/A
	Current Research	Tetra	GAAA	3.985	322.7-438.7	29 (19723594-19723782)
FH3313	Guyon et al. (2003)	Tetra	Motif not defined	N/A	N/A	N/A
	Current Research	Tetra	GAAA	3.879	340.93-445.69	19 (24606038-24606459)
FH3377	Guyon et al. (2003)	Penta	Motif not defined	N/A	N/A	N/A

	Current Research	Penta	GAAAA	4.675	183.01-305.21	3 (78748898-78749090)
PEZ02	Eichmann et al. (2004a; b)	Tetra	GGAA	N/A	N/A	N/A
	Current Research	Tetra	GGAA	4.011	104.36-144.36	17 (13276076-13276209)
PEZ05	Halverson and Basten (2005)/ Halverson et al. (17)	Tetra	AAAG	N/A	N/A	N/A
	Current Research	*Tetra*	*TTTA*	*3.967*	*92.48-116.24*	*12 (60326434-60326541)*
PEZ16	Halverson and Basten (1)/ Halverson et al. (1999)	Tetra	AAAG	N/A	N/A	N/A
	Current Research	*Tetra*	*GAAA*	*3.935*	*280.7-331.66*	*27 (10305692-10305995)*
PEZ17	Halverson and Basten (2005)/ Halverson et al. (1999)	Tetra	AAAG	N/A	N/A	N/A
	Current Research	*Tetra*	*GAAA*	*4.225*	*190.98-224.58*	*4 (71904833-71905038)*
PEZ21	Halverson and Basten (2005)/ Halverson et al. (1999)	Tetra	AAAT	N/A	N/A	N/A
	Current Research	Tetra	AAAT	4.015	83.02-103.22	2 (36438658-36438751)
VWF.X	Shibuya et al. (1994)	Hexa	AGGAAT	N/A	N/A	N/A
	Current Research	Hexa	AGGAAT	5.965	151.1-186.74	27 (41977918-41978074)

Table 2 Annealing temperatures for each locus.

Locus	Annealing Temp. °C
FH2001	51
FH2004	64
FH2010	57
FH2017	58
FH2054	57
FH2088	56
FH2107	54
FH2309	52
FH2328	58
FH2361	59
FH3313	57
FH3377	54
PEZ02	60
PEZ05	57
PEZ16	57
PEZ17	59
PEZ21	52
vWF.X	57
ZFx/ZFy	57

Table 3 Observed national and regional STR allele frequencies (n = number of different allele types).

Locus	Binned allele classes (bp)		Regional				National=Total (N=667)
	Length[1]	Mean[2]	Western (N=147)	Southern (N=241)	Mid-western (N=164)	Northeastern (N=115)	
FH2001	118.77	118.803	0.0068	0.0062	0.0152	0.0217	0.0112
n=11	122.89	122.736	0.0204	0.027	0.0122	0.013	0.0195
	127.01	126.866	0.2755	0.2137	0.2287	0.2304	0.2339
	131.13	130.992	0.0646	0.0622	0.061	0.0478	0.06
	135.25	135.099	0.034	0.0581	0.0701	0.0696	0.0577
	139.37	139.283	0.3197	0.2386	0.3079	0.313	0.2864
	143.49	143.724	0.2109	0.2697	0.1768	0.1957	0.2211
	147.61	148.187	0.051	0.0622	0.0671	0.087	0.0652
	151.73	152.549	0.0102	0.0228	0.0152	0.013	0.0165
	155.85	154.552	0.0068	0.0311	0.0396	0	0.0225
	159.97	158.648	0	0.0083	0.0061	0.0087	0.006
FH2004	232.82	232.778	0.1531	0.112	0.128	0.1522	0.1319
n=17	237.02	237.028	0.3231	0.3444	0.3262	0.2739	0.3231
	241.22	241.256	0.2755	0.2863	0.2713	0.2783	0.2789
	245.42	245.446	0.0816	0.139	0.1311	0.1696	0.1297
	249.62	249.578	0.0136	0.0124	0.0091	0.0174	0.0127
	279.02	278.603	0	0.0041	0.003	0	0.0022
	283.22	284.41	0	0	0.003	0.0087	0.0022
	287.42	286.366	0.017	0.0021	0.0091	0	0.0067
	291.62	290.21	0.0034	0	0	0	0.0007
	295.82	296.033	0.0034	0.0021	0	0.0043	0.0022
	300.02	299.892	0.0306	0.027	0.0305	0.0261	0.0285
	304.22	304.114	0.017	0.0228	0.0183	0.0304	0.0217
	308.42	308.3	0.0136	0.0021	0.0061	0.0087	0.0067
	312.62	312.663	0.017	0.0145	0.0152	0.0087	0.0142
	316.82	316.971	0.0102	0.0104	0.0183	0.0043	0.0112
	321.02	321.097	0.0374	0.0207	0.0305	0.013	0.0255
	325.22	325.245	0.0034	0	0	0.0043	0.0015
FH2010	221.66	221.66	0	0	0	0.0043	0.0007
n=6	225.86	225.821	0.085	0.1183	0.1585	0.1609	0.1282
	230.06	230.037	0.3265	0.4149	0.3293	0.3696	0.3666
	234.26	234.204	0.3299	0.2199	0.253	0.1826	0.2459

[1] The expected lengths (ELs) of alleles placed in the bin were estimated using Amos et al.'s (20) methods. ELs are not observed lengths (OLs). OLs are based on the actual mobility rates of amplicons and are raw and unbinned data.
[2] Mean lengths of alleles placed in the bin

	238.46	238.384	0.2415	0.2448	0.2561	0.2739	0.2519
	242.66	242.554	0.017	0.0021	0.003	0.0087	0.0067
FH2017	256.69	256.586	0.017	0.0477	0.0488	0.0609	0.0435
n=6	260.49	260.494	0.034	0.0477	0.061	0.0261	0.0442
	264.29	264.312	0.6667	0.6245	0.6189	0.5609	0.6214
	268.09	268.177	0.2075	0.2407	0.2195	0.313	0.2406
	271.89	271.999	0.068	0.0394	0.0488	0.0391	0.048
	275.69	275.773	0.0068	0	0.003	0	0.0022
FH2054	139.09	138.508	0.0034	0.0083	0	0	0.0037
n=10	143.25	143.028	0.068	0.0643	0.064	0.0304	0.0592
	147.41	147.427	0.1973	0.1971	0.2256	0.2522	0.2136
	151.57	151.776	0.2211	0.2759	0.2652	0.2304	0.2534
	155.73	155.986	0.0952	0.0871	0.0945	0.1087	0.0945
	159.89	160.096	0.0816	0.0809	0.0884	0.0783	0.0825
	164.05	164.115	0.1939	0.1452	0.1006	0.1348	0.1432
	168.21	168.139	0.0748	0.0934	0.1037	0.0913	0.0915
	172.37	172.156	0.0612	0.0373	0.0518	0.0696	0.0517
	176.53	176.23	0.0034	0.0104	0.0061	0.0043	0.0067
FH2088	94.56	95.49	0.0068	0.0021	0	0	0.0022
n=8	110.4	110.709	0.0272	0.0249	0.0366	0.0261	0.0285
	118.32	118.349	0.2075	0.2095	0.2409	0.2	0.2151
	122.28	122.253	0.0714	0.0685	0.0884	0.0565	0.072
	126.24	126.233	0.2755	0.3154	0.2988	0.2652	0.2939
	130.2	130.346	0.2279	0.2344	0.2195	0.3478	0.2489
	134.16	134.312	0.1837	0.1432	0.1159	0.1043	0.1387
	138.12	138.4	0	0.0021	0	0	0.0007
FH2107	291.72	292.68	0	0	0	0.0043	0.0007
n=22	295.44	296.38	0	0	0	0.0043	0.0007
	347.52	347.18	0	0	0.003	0	0.0007
	351.24	350.835	0	0.0041	0	0	0.0015
	354.96	354.565	0.0068	0	0	0	0.0015
	358.68	358.453	0.0136	0.0228	0.0366	0.0261	0.0247
	362.4	362.256	0.0374	0.0539	0.0732	0.0391	0.0525
	366.12	365.988	0.0476	0.0622	0.0396	0.0696	0.0547
	369.84	369.769	0.119	0.1535	0.1799	0.1522	0.1522
	373.56	373.588	0.2075	0.1805	0.186	0.1957	0.1904
	377.28	377.386	0.2347	0.1888	0.1555	0.1217	0.1792
	381	381.175	0.1395	0.1515	0.1372	0.1435	0.1439
	384.72	384.969	0.0986	0.0581	0.0732	0.1	0.078

	388.44	388.792	0.0374	0.0415	0.0366	0.0478	0.0405
	392.16	392.394	0.0102	0.0145	0.0213	0.013	0.015
	395.88	395.364	0.0204	0.029	0.0274	0.0348	0.0277
	399.6	398.567	0.0068	0.0145	0.0091	0.0261	0.0135
	403.32	402.416	0.0136	0.0083	0.0183	0.0217	0.0142
	407.04	406.508	0.0034	0.0104	0	0	0.0045
	410.76	410.457	0	0.0041	0.003	0	0.0022
	418.2	418.01	0.0034	0	0	0	0.0007
	425.64	425.85	0	0.0021	0	0	0.0007
FH2309	339.66	339.882	0.0374	0.027	0.0183	0.0391	0.0292
n=22	343.5	343.794	0.0374	0.0768	0.0793	0.0478	0.0637
	347.34	347.613	0.0068	0.029	0.0488	0.0348	0.03
	351.18	351.448	0.0408	0.0062	0.0061	0.0043	0.0135
	355.02	355.488	0.0238	0.027	0.0366	0.0478	0.0322
	358.86	359.417	0.0034	0.0062	0.0061	0.0043	0.0052
	362.7	363.379	0.0068	0.0062	0.003	0.0087	0.006
	366.54	365.805	0.0238	0.0145	0.0152	0.0217	0.018
	370.38	369.731	0.0408	0.0394	0.0549	0.0391	0.0435
	374.22	374.345	0.0306	0.0207	0.0183	0.0522	0.0277
	378.06	377.418	0.0374	0.0207	0.0213	0.0217	0.0247
	381.9	381.335	0.0782	0.0788	0.1037	0.0696	0.0832
	385.74	385.271	0.0442	0.0768	0.1067	0.0609	0.0742
	389.58	389.252	0.1156	0.1141	0.0732	0.0696	0.0967
	393.42	393.247	0.1497	0.1349	0.1006	0.1304	0.1289
	397.26	397.27	0.1259	0.1432	0.1037	0.1174	0.1252
	401.1	401.332	0.0306	0.0498	0.0305	0.0217	0.036
	404.94	405.356	0.085	0.0332	0.0488	0.0435	0.0502
	408.78	409.345	0.0578	0.0685	0.0854	0.113	0.078
	412.62	413.395	0.0204	0.0124	0.0274	0.0478	0.024
	416.46	417.473	0	0.0145	0.0122	0.0043	0.009
	427.98	429.5	0.0034	0	0	0	0.0007
FH2328	171	170.92	0	0.0021	0	0	0.0007
n=12	174.84	174.81	0.034	0.0622	0.0762	0.1043	0.0667
	178.68	178.618	0.0068	0.0041	0	0.0087	0.0045
	182.52	182.519	0.1667	0.1535	0.186	0.2	0.1724
	186.36	186.378	0.1429	0.2178	0.1555	0.1435	0.1732
	190.2	190.215	0.1905	0.1929	0.1494	0.1174	0.1687
	194.04	194.019	0.2041	0.1203	0.1646	0.1304	0.1514
	197.88	197.865	0.1565	0.139	0.1463	0.1696	0.1499
	201.72	201.752	0.0578	0.0726	0.0915	0.0826	0.0757
	205.56	205.805	0.0306	0.0311	0.0244	0.0261	0.0285

	209.4	209.848	0.0034	0.0021	0.003	0.0087	0.0037
	213.24	213.968	0.0068	0.0021	0.003	0.0087	0.0045
FH2361	322.7	322.4	0	0.0021	0.003	0.0043	0.0022
n=29	326.7	326	0	0.0021	0	0	0.0007
	330.7	330.498	0.0136	0.0041	0.0061	0.0043	0.0067
	334.7	334.61	0.0204	0.0373	0.0366	0.0391	0.0337
	338.7	338.872	0.0918	0.0809	0.0762	0.0609	0.0787
	342.7	342.812	0.2245	0.2469	0.2378	0.2043	0.2324
	346.7	346.609	0.3095	0.2448	0.2439	0.2348	0.2571
	350.7	350.581	0.1735	0.1515	0.1677	0.2217	0.1724
	354.7	354.531	0.0816	0.0871	0.0823	0.1087	0.0885
	358.7	358.511	0.017	0.0436	0.0427	0.0304	0.0352
	362.7	362.541	0.0204	0.0332	0.0366	0.0304	0.0307
	366.7	366.783	0.0034	0.0041	0.0061	0.0087	0.0052
	370.7	370.743	0.0068	0.0124	0.0061	0	0.0075
	374.7	374.133	0.0034	0.0021	0	0.0043	0.0022
	378.7	378.103	0.0034	0.0041	0	0	0.0022
	382.7	382.21	0	0.0021	0	0	0.0007
	390.7	390.17	0.0034	0.0021	0	0	0.0015
	394.7	394.15	0	0	0	0.0043	0.0007
	398.7	398.16	0	0	0.003	0	0.0007
	402.7	402.28	0	0.0021	0.003	0	0.0015
	406.7	406.636	0.0034	0.0062	0.003	0	0.0037
	410.7	410.385	0.0136	0.0083	0.003	0.013	0.009
	414.7	414.422	0.0068	0.0083	0.0061	0.013	0.0082
	418.7	418.547	0	0.0041	0.0122	0	0.0045
	422.7	422.524	0	0.0041	0.0091	0.0087	0.0052
	426.7	426.588	0	0.0041	0.0061	0	0.003
	430.7	430.605	0	0.0021	0.0061	0.0043	0.003
	434.7	434.745	0	0	0.003	0.0043	0.0015
	438.7	438.78	0.0034	0	0	0	0.0007
FH3313	340.93	340.37	0	0	0.003	0	0.0007
n=25	352.57	352.42	0.0034	0	0	0	0.0007
	360.33	360.057	0	0.0041	0	0.0043	0.0022
	364.21	363.869	0.0034	0.0166	0.0122	0.013	0.012
	368.09	368.274	0.0204	0.0207	0.0152	0.0304	0.021
	371.97	372.26	0.0442	0.0705	0.0366	0.0565	0.054
	375.85	376.013	0.0646	0.0436	0.0274	0.0478	0.045
	379.73	379.857	0.034	0.0602	0.0671	0.0652	0.057
	383.61	383.771	0.0476	0.0747	0.0915	0.0522	0.069
	387.49	387.468	0.0272	0.0436	0.0366	0.0435	0.0382

	391.37	391.145	0.0306	0.0332	0.0457	0.0522	0.039
	395.25	395.093	0.0442	0.0622	0.0549	0.0652	0.057
	399.13	398.82	0.0612	0.0581	0.0427	0.0565	0.0547
	403.01	402.448	0.0374	0.0581	0.0488	0.0261	0.0457
	406.89	406.437	0.1054	0.0519	0.0457	0.0435	0.0607
	410.77	410.27	0.0782	0.0726	0.0823	0.0696	0.0757
	414.65	414.337	0.0884	0.0871	0.1098	0.1304	0.1004
	418.53	418.408	0.1429	0.1017	0.1463	0.1261	0.1259
	422.41	422.308	0.0612	0.0394	0.0549	0.0652	0.0525
	426.29	426.649	0.0238	0.0332	0.0244	0.0087	0.0247
	430.17	430.99	0.034	0.0166	0.0274	0.013	0.0225
	434.05	434.627	0.0238	0.0145	0.0122	0.0174	0.0165
	437.93	438.401	0.0068	0.0228	0.0061	0.0087	0.0127
	441.81	442.972	0.0136	0.0104	0.0061	0.0043	0.009
	445.69	446.873	0.0034	0.0041	0.003	0	0.003
FH3377	183.01	182.839	0.0306	0.0851	0.1006	0.087	0.0772
n=22	187.71	187.664	0.068	0.0913	0.0915	0.0783	0.084
	192.41	192.456	0.1667	0.1515	0.1402	0.1174	0.1462
	197.11	197.292	0.2007	0.1515	0.189	0.2435	0.1874
	201.81	202.144	0.0748	0.0996	0.0671	0.0957	0.0855
	206.51	207.299	0.0612	0.0456	0.0457	0.0348	0.0472
	211.21	212.357	0.017	0.0083	0.0152	0.0217	0.0142
	230.01	228.578	0	0.0041	0	0.0087	0.003
	234.71	233.721	0.0646	0.0311	0.0305	0.0174	0.036
	239.41	238.787	0.1497	0.1224	0.128	0.1217	0.1297
	244.11	243.936	0.0918	0.0934	0.0915	0.0478	0.0847
	248.81	248.975	0.0578	0.056	0.0579	0.0826	0.0615
	253.51	253.535	0	0.0166	0.0122	0.0087	0.0105
	262.91	263.58	0	0.0021	0.003	0	0.0015
	267.61	267.568	0.0034	0.0062	0.0061	0	0.0045
	272.31	272.82	0.0034	0.0041	0	0.0043	0.003
	281.71	281.819	0	0.0062	0.0122	0.0087	0.0067
	286.41	286.812	0.0068	0.0083	0.0061	0.0087	0.0075
	291.11	291.637	0	0.0104	0.003	0.0043	0.0052
	295.81	296.43	0	0.0021	0	0.0043	0.0015
	300.51	298.313	0	0.0041	0	0.0043	0.0022
	305.21	303.09	0.0034	0	0	0	0.0007
VWF.X	151.1	150.77	0	0.0041	0.003	0.0043	0.003
n=7	157.04	156.992	0.4932	0.473	0.439	0.5609	0.4843
	162.98	163.063	0.3707	0.3714	0.4329	0.2913	0.3726
	168.92	169.044	0.0646	0.0768	0.0671	0.0913	0.0742

	174.86	174.994	0.0476	0.0228	0.0091	0.0217	0.0247
	180.8	180.965	0.0068	0.029	0.0244	0.0087	0.0195
	186.74	186.856	0.017	0.0228	0.0244	0.0217	0.0217
PEZ02	104.36	104.479	0.017	0.0207	0.0427	0.0043	0.0225
n=11	108.36	108.397	0.0034	0.0041	0	0	0.0022
	112.36	112.351	0.0306	0.0539	0.0213	0.0652	0.0427
	116.36	116.301	0.0578	0.0768	0.1159	0.113	0.0885
	120.36	120.308	0.1973	0.2095	0.2165	0.1826	0.2039
	124.36	124.309	0.3401	0.334	0.3018	0.313	0.3238
	128.36	128.352	0.2721	0.2552	0.2622	0.2783	0.2646
	132.36	132.448	0.0714	0.029	0.0335	0.0304	0.0397
	136.36	136.531	0.0034	0.0124	0.003	0.013	0.0082
	140.36	140.8	0	0	0.003	0	0.0007
	144.36	145.188	0.0068	0.0041	0	0	0.003
PEZ05	92.48	92.0889	0.0034	0.0166	0.0244	0.0087	0.0142
n=7	96.44	96.1071	0	0.0104	0.003	0.0043	0.0052
	100.4	100.366	0.466	0.4191	0.4421	0.5217	0.4528
	104.36	104.3	0.1531	0.1556	0.1677	0.1435	0.1559
	108.32	108.248	0.2823	0.2863	0.2744	0.287	0.2826
	112.28	112.255	0.0952	0.112	0.0884	0.0261	0.0877
	116.24	116.2	0	0	0	0.0087	0.0015
PEZ16	280.7	281.123	0.0034	0.0021	0	0.0043	0.0022
n=14	284.62	284.868	0.0578	0.056	0.0335	0.0391	0.048
	288.54	288.893	0.102	0.0788	0.0457	0.0783	0.0757
	292.46	292.603	0.1497	0.1971	0.2256	0.1609	0.1874
	296.38	296.189	0.1395	0.1266	0.1159	0.1652	0.1334
	300.3	299.953	0.2925	0.2967	0.3567	0.3348	0.3171
	304.22	304.258	0.1361	0.0913	0.1159	0.1	0.1087
	308.14	308.456	0.0544	0.0415	0.0305	0.0435	0.042
	312.06	312.392	0.0102	0.0124	0	0.0174	0.0097
	315.98	316.541	0.034	0.0685	0.0518	0.0435	0.0525
	319.9	319.9	0	0.0041	0.0091	0	0.0037
	323.82	324.312	0.017	0.0104	0.0122	0.0043	0.0112
	327.74	328.169	0.0034	0.0104	0.003	0.0043	0.006
	331.66	332.59	0	0.0041	0	0.0043	0.0022
PEZ17	190.98	191.38	0	0	0.003	0	0.0007
n=9	195.18	195.262	0.0306	0.0311	0.0305	0.0478	0.0337
	199.38	199.263	0.1701	0.1784	0.2043	0.187	0.1844
	203.58	203.486	0.2347	0.2365	0.1707	0.2348	0.2196

	207.78	207.757	0.2449	0.2344	0.2561	0.2217	0.2399
	211.98	211.995	0.2279	0.222	0.2073	0.2304	0.2211
	216.18	216.225	0.0816	0.0768	0.0945	0.0565	0.0787
	220.38	220.465	0.0102	0.0145	0.0335	0.0217	0.0195
	224.58	224.437	0	0.0062	0	0	0.0022
PEZ21	83.02	83.1112	0.0374	0.0643	0.064	0.0478	0.0555
n=6	87.06	87.0828	0.3605	0.3008	0.3079	0.2826	0.3126
	91.1	91.0312	0.1497	0.1577	0.1037	0.1391	0.1394
	95.14	95.0956	0.415	0.4087	0.4878	0.4826	0.4423
	99.18	99.1534	0.0374	0.0622	0.0366	0.0478	0.048
	103.22	103.05	0	0.0062	0	0	0.0022

Table 4 Peak sizes and RFU values for peaks at the canine 19-plex loci for 13 single hair extraction samples (samples 1-13) and 4 saliva stain extractions (samples 14-17). Next to the marker name, alleles are described in raw base lengths and are followed by a backslash (/) and the height of the peak is listed in RFUs.

Sample	Locus	Peak 1 (bp)/RFU	Peak 2 (bp)/RFU	Peak 3 (bp)/RFU
1	Zinc. Finger	159.31/70	163.77/55	
2	-	-	-	
3	-	-	-	
4	PEZ02	125.86/28	-	
	Zinc. Finger	159.21/67	-	
	FH2328	190.35/55	-	
	FH2004	239.67/27	-	
	FH2361	350.13/28	-	
5	-	-	-	
6	Zinc. Finger	159.21/67	-	
	PEZ17	207.63/50	-	
	FH3377	192.70/30	-	
7	PEZ02	125.83/46	-	
	Zinc. Finger	159.22/67	163.67/32	
	PEZ05	101.09/59	-	
	FH2001	148.92/79	-	
	FH2361	354.23/77	-	
8	PEZ02	125.76/54	129.87/50	
	Zinc. Finger	159.29/173	163.75/50	
	PEZ17[3]	199.39/45	203.57/49	215.79/47
	FH2309	355.04/92	-	
	PEZ05	101.07/234	-	
	FH2001	140.71/155	148.92/132	
	FH2328	182.46/87	193.96/60	
	FH2004	239.60/73	-	
	FH2361	346.16/57	350.24/92	
	PEZ21	87.39/102	-	
	FH2054	152.20/44	164.34/70	
	FH3377	197.38/46	246.90/40	
	FH2107	380.37/54	-	
	FH2088	131.89/224	-	
	vWF.X	157.09/115	-	
	FH2010	236.78/120	-	
	PEZ16	304.25/112	-	
	FH3313	402.70/53	414.37/55	
9	PEZ05	101.12/50	-	
10	Zinc. Finger	159.14/47	163.72/36	
	PEZ05	105.11/35	-	
	FH2001	128.34/30	144.61/34	
	FH2328	197.91/38	-	

	FH2361	354.15/31	-	
	PEZ21	87.30/27	-	
	FH2054	168.43/37	172.34/31	
	FH3377	246.56/46	-	
	FH2088	123.49/50	131.75/35	
	FH3313	383.98/51	-	
11	Zinc. Finger	159.15/25	-	
	PEZ05	101.01/53	-	
	FH2001	144.80/56	148.87/26	
	FH2328	182.47/30	-	
	FH2361	345.97/25	-	
	PEZ21	87.41/49	-	
	FH2054	152.09/39	-	
	FH2088	131.75/49	-	
12	Zinc. Finger	159.16/38	-	
	FH2001	148.90/37	-	
	FH2004	239.43/53	-	
	FH2361	346.12/31	-	
	FH2054	164.36/43	-	
	FH3377	197.41/53	-	
	FH2010	236.69/46	-	
13	PEZ02	121.74/40	-	
	Zinc. Finger	159.08/62	-	
	PEZ17	203.55/68	207.53/35	215.73/35
	FH2017	266.77/86	-	
	FH2001	140.70/31	-	
	FH2361	350.06/30	-	
	PEZ21	87.37/52	-	
	FH2054	152.12/28	164.32/64	
	FH2010	236.70/42	-	
	PEZ16	288.36/27	299.70/43	
	FH3313	402.71/48	-	
14	PEZ02	130.00/45	-	
	Zinc. Finger	159.23/53	163.70/73	
	PEZ17	203.50/137	215.73/64	
	PEZ05	101.18/137	-	
	FH2001	140.71/101	148.86/131	
	FH2328	182.50/69	193.87/93	
	FH2004	239.49/79	-	
	PEZ21	83.39/83	87.42/49	
	FH2054	152.21/63	164.29/34	
	FH3377	187.92/44	197.44/44	
	FH2088	131.81/97	-	
15	-	-	-	
16	-	-	-	
17	PEZ02	129.91/64	-	
	Zinc. Finger	159.23/113	163.77/62	
	PEZ17	215.57/92	-	
	PEZ05	101.08/126	-	

	FH2001	140.75/73	-	
	FH2328	193.94/29	-	
	FH2004	239.62/68	-	
	PEZ21	83.40/64	87.49/54	
	FH2054	152.15/187	-	
	FH3377	197.44/95	-	
	FH2088	131.93/199	-	
	vWF.X	157.10/94	-	
	FH2010	229.02/55	-	
	PEZ16	288.36/54	-	

*** PEZ17 had three allele calls in two different samples. The low RFU values make it difficult to determine which allele calls are true peaks.

Table 5 Allele calls and frequencies for the Finnzymes multiplex kit's positive control animal (F-863) and for Non-probative Sample 8. 'Raw' and 'Binned' values are in bases.

Locus	**Positive Control**		**Non-probative sample**		National
	Raw	Binned	Raw	Binned	Freq.
FH2001	128.44	127.01	-	-	0.2339
	-	-	140.71	139.37	0.2864
	144.76	143.49	-	-	0.2211
	-	-	148.92	147.61	0.0652
FH2004	231.51	232.82	-	-	0.1319
	239.63	241.22	239.6	241.22	0.2789
FH2010	232.79	234.26	-	-	0.2459
	-	-	236.78	238.46	0.2519
FH2017	262.64	264.29	-	-	0.6214
	266.76	268.09	-	-	0.2406
FH2054	148.08	147.41	-	-	0.2136
	-	-	164.34	164.05	0.1432
	168.26	168.21	-	-	0.0915
FH2088	123.65	122.28	-	-	0.072
	127.58	126.24	-	-	0.2939
	-	-	131.89	130.2	0.2489
FH2107	369.77	369.84	-	-	0.1522
	-	-	380.37	381	0.1439
	389.23	388.44	-	-	0.0405
FH2309	-	-	355.04	355.02	0.0322
	393.9	393.42	-	-	0.1289
FH2328	169.79	171			0.0007
	-	-	182.46	182.52	0.1724
	-	-	193.96	194.04	0.1514
	205.47	205.56	-	-	0.0285
FH2361	342.1	342.7	-	-	0.2324
	343.99	342.7	-	-	0.2324
	-	-	346.16	346.7	0.2571
	-	-	350.24	350.7	0.1724
FH3313	-	-	402.7	403.01	0.0457
	-	-	414.37	414.65	0.1004
	418.21	418.53	-	-	0.1259
	423.84	422.41	-	-	0.0525
FH3377	197.4	197.11	-	-	0.1874
vWF.X	157.06	157.04	157.09	157.04	0.4843
PEZ02	-	-	125.76	124.36	0.3238
	129.8	128.36	129.87	128.36	0.2646
PEZ05	101.04	100.4	101.07	100.4	0.4528
PEZ16	300.32	300.3	-	-	0.3171
	304.19	304.22	304.25	304.22	0.1087
PEZ17	199.47	199.38	-	-	0.1844
	211.56	211.98	-	-	0.2211
PEZ21	87.44	87.06	87.39	87.06	0.3126

	95.68	95.14	-	-	0.4423

Figure 1 Geographic locations in the U.S. that are represented by the samples used in this study.

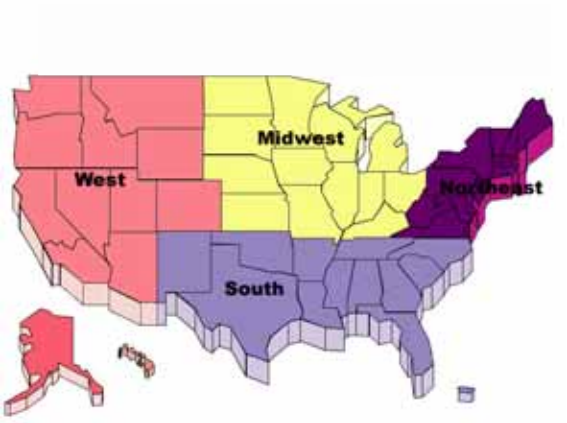


Figure 2 The positive control F-863 STR profile and the allelic distribution at each of the 19 loci.

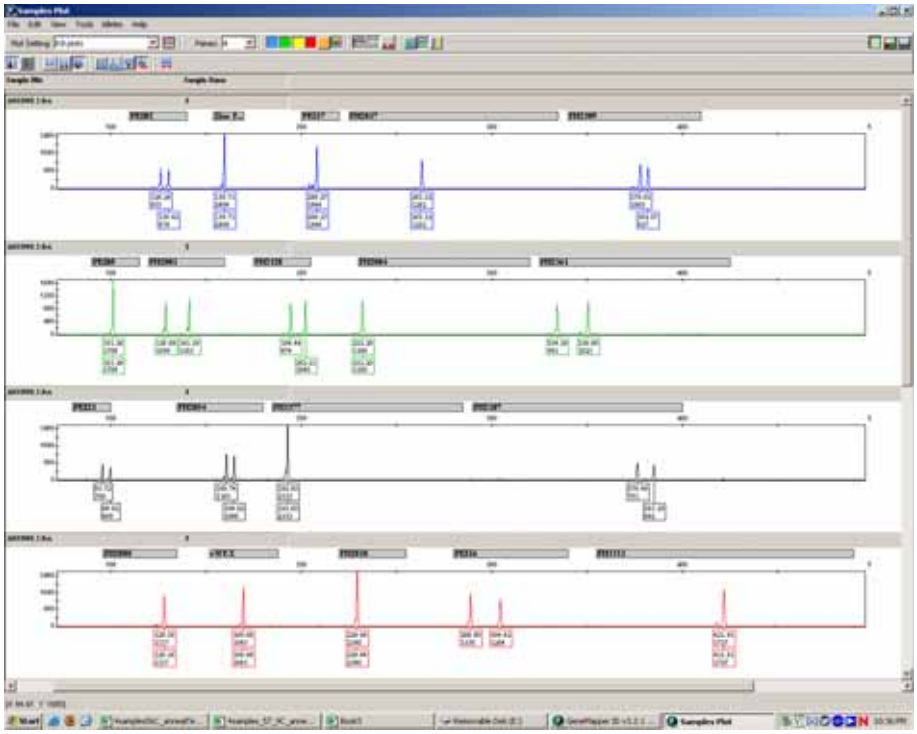


Figure 3 Assignment results of the STRUCTURE analysis based on the pedigreed breed and mixed breed dog samples, where K = 13 breeds and 1. American Pit Bull, 2. Beagle, 3. Dachshund, 4. German Shepherd, 5. Golden Retriever, 6. Labrador Retriever, 7. Miniature

Poodle, 8. Standard Poodle, 9. Rottweiler, 10. Shih Tzu, 11. Toy Poodle, 12. Yorkshire Terrier, and 13. mixed breed dogs.

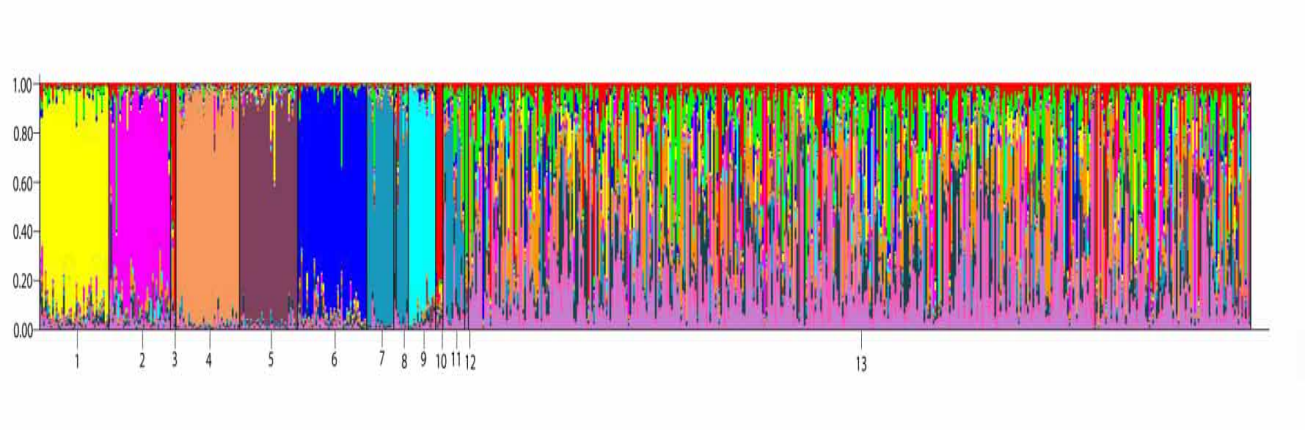
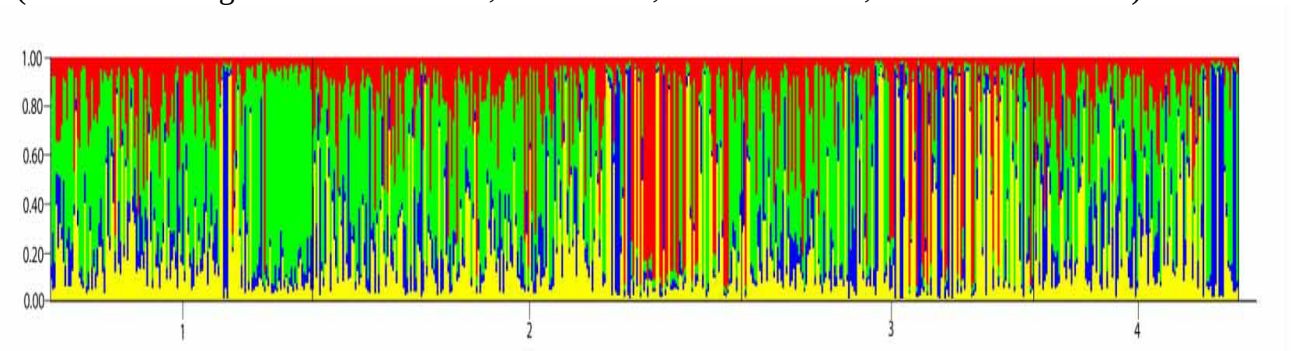


Figure 4 Assignment results of the STRUCTURE analysis based on the four regional samples (where K = 4 regions and 1. Western, 2. Southern, 3. Mid-western, and 4. Northeastern)



IV. Conclusions

With approximately 70 million dogs residing in households in the United States, canine samples (whether hair, saliva from bites, or others) are often part of the physical evidence associated with crime scenes. Until now there were no canine testing kits that have been developed and peer-reviewed, standardized, and validated for conducting forensic analysis of canine biological specimens. This NIJ funded project focused on the development and validation of a canine multiplex reagent kit for use and the direct goals of the project that was proposed to NIJ.

The population study was been carried out using 18 STR loci selected specifically for identity testing of canines. To our knowledge this is the first investigation of STR diversity and genetic subdivision among dogs of pedigreed and mixed ancestry across different regions of the U.S. The SWGDAM-like development validation study was akin to validation tests conducted on human identification kits for the analysis of challenged samples. A nomenclature for these canine-specific markers included was also developed based on internationally recognized recommendations for human forensic STR loci.

By establishing and validating a panel of canine STRs that is *designed for forensic application*, we can address those issues raised by the courts and add the powerful tool of canine DNA typing to the forensic science arsenal.

Population study

Regardless of breed, mixed breed, or geographic region, the gene diversity for the combined 18 loci is high. Therefore, it can be anticipated that these loci will be useful for

identity testing for most forensic and kinship analyses. However, because of the known selection and inbreeding history of the domestic dog imposed by man, it is important to assess the impact of population substructure and how it may impact estimates of the rarity of a canine STR profile. On average, only 3 locus pairs per breed across the 12 pedigreed breeds and one population of mixed breed dogs tested were found to be out of linkage equilibrium at the $p = 0.01$ level. The number of pairs of loci out of equilibrium was greater among mixed breed than pedigreed dogs. This observation concurs with that of Halverson and Basten's (2005) which was based on private sample collections including samples from the American Kennel Club (AKC). Even pedigreed German Shepherds, which have the lowest diversity among all categories of dogs, did not show any substantial linkage disequilibrium within and between loci. Estimates of LD between loci based on the current data set were not statistically significant for any syntenic loci. This could be because the relative distance between these loci on their respective chromosomes ranged from 5×10^6 bp to 32×10^6 bp (with recombination factors of 0.05 to 0.32) giving alleles at these loci ample opportunity to segregate independently. While greater sample numbers per breed would be desirable, the data are consistent with other studies and support that allele frequencies across all loci analyzed here can be used to calculate random match, parentage exclusion and breed assignment probabilities.

The mean observed and expected heterozygosity estimates reported here, especially those for Golden and Labrador Retrievers, are higher than comparable measures reported by Irion et al. (2003) and DeNise et al. (2004). This difference, in combination with the within-locus Hardy-Weinberg (HW) disequilibria among three of the most outbred dog breeds (i.e., the American Pit Bull, Golden retriever and the mixed breed dogs), implies that inbreeding is not the only factor that has shaped the genetic structure of the domestic dog. While our genetic diversity estimates are comparable to those estimated by Halverson and Basten (2005), some cross-study differences are evident. For example, the pedigreed Dachshunds exhibited the highest degree of genetic diversity among all dogs (although a sample of only 3 animals could have biased our estimates), and Pit Bulls were more genetically diverse than the combined breeds of Poodles in this study. Halverson and Basten (2005), however, also ranked the Dachshunds highest in genetic diversity after the Poodles, followed by Yorkshire Terriers, then Pit Bulls. The Dachshunds also exhibit the most divergent Y chromosome haplotypes among American, Asian, Australian and European dog breeds (32). There are several varieties of Dachshunds based on size, coat type, and color; therefore, the wide assortment of morphometric differences concur with the high level of genetic diversity within this breed.

While the analysis of regionally diverse samples in our study could have inflated our heterozygosity estimates, Fst measurements show that regional variation contributed only 0.2% of the genetic differences among U.S. dog populations (based on geography). Furthermore, the PCA and STRUCTURE analyses of regional populations are consistent with the outcome that there is little genetic differentiation among groups of mixed-breed dogs originating from different geographical regions within the U.S. Regardless of whether the analysis included *a priori* defined geographic groups, when up to four regional populations of domestic dogs were assumed (i.e., $K = 2 – 4$), no distinct STR distributions emerged with regard to the four geographic regions.

While the population substructure among regional populations is very small, there is a mild correlation between the amount of genetic distance and geographic distance. Genetic relationships among dog populations correlate sllightly with geographic transition in the U.S. As an example, dogs from the western and northeastern states are more distantly related compared to dogs from the latter region and the mid-western states.

Even though STR loci tend to have relatively high mutation rates, regional Fst values show extremely low levels of variation among regions and suggest sufficient amounts of gene flow among regions of the U.S. to reduce significant genetic subdivision through genetic drift. Such low levels of regional variation are consistent with the hypothesis that pet-owners take their

pets with them when they migrate. Each regional sample also represented the genetic variation of the much broader nationwide sample, which confirms findings of Himmelberger et al. (2008) and Baute et al. (2008) that were based on canine mtDNA. As such, in forensic casework, estimates of genetic diversity based on regional populations proximate to the crime scene may also exhibit a similar genetic structure represented by samples collected from different and diverse geographic regions. Conversely, in the absence of a local STR database and no knowledge of the breed or breed make-up of the donor of the biological evidence, investigators can rely on a global database for estimating canine DNA match probabilities.

Previous studies have attributed the wide genetic variation that exists among current dog breeds to each breed's unique breeding history and country of origin (Halverson and Basten 2005; Irion et al. 2003). Among-breed Fst values of 0.09 imply that genetic divergence among domestic dog populations in the U.S. is moderate (although much higher than that for humans – about 10 times the recommended pragmatic value [NRC 1996]). This among-breed Fst value is comparable to the estimate computed by Halverson and Basten (2005). Since 9% of variation is attributed to genetic differences among the various dog breeds, approximately 91% of the genetic diversity is found within breed types. While Fst is influenced by the effective size and degree of gene flow among dog populations, most of this variation, especially that among pedigreed dogs and between pedigreed and mixed breed dogs, probably results from strong genetic drift due to small effective population sizes resulting from the artificial selection practices carried out in the domestic dog.

The positive Fis values, which are in agreement with the gene diversity estimates among breeds and across regions, indicate that there is an increased number of homozygotes in almost all populations of pedigreed dog breeds reflecting the extent of genetic isolation (or inbreeding) still extant among these, if not all, dog breeds. Results based on pedigreed dogs, which include the reduced number of loci that are out of equilibrium among pedigreed dogs compared to mixed and pure breed dogs, are concordant with the breeding strategies of kennel clubs to outcross pedigreed dogs in order to maximize their genetic diversity while still maintaining the rigid genetic boundaries that preserve individual breed standards. German Shepherds, despite being the most highly derived breed based on low estimates of observed and expected heterozygosity, exhibited more heterozygous individuals than would be expected in an inbred breed.

In the STRUCTURE and GeneClass analyses, the dichotomy between pedigreed and mixed breed dogs became obvious since each of the pedigreed breeds clustered tightly as distinct genetic groups and retaining the original assignment probabilities. This is consistent with STR data for humans where the high diversity of the forensically-selected loci provides a high power of discrimination, but the populations could be separated according to their known ethnohistory (Budowle 2003). Our STR results are different than those of Himmelberger et al.'s (2008) mtDNA-based results which concluded that no significant variation exists between population structures of pedigreed and mixed breed dogs. The different dog breeds studied here were distributed evenly across the four regions and is consistent with AKC's survey (http://www.akc.org/) of top dog breed distributions across U.S. states. Regional allele frequency distribution is very similar even though breed-specific allele frequencies were significantly different. Interestingly, the different Poodle types cluster into a common group despite their morphological differences in size and shape with Toy Poodles exhibiting the highest degree of genetic introgression from other breeds. Toy Poodles had the most variable gene pool and fewest breed-specific STR alleles than any other breed including the other poodle-type dogs. The Toy Poodle's genetic composition reflects a much more recent and more complex breed development history than the other Poodle subtypes.

The U.S. Center for Disease Control (CDC) has attributed most dog-bite fatalities in humans to Pit Bulls and Rottweilers (Gershman et al. 1994; Sacks et al. 1996). Therefore, these two breeds are increasingly involved in litigation in which genetic testing can contribute. We tested the 18 STRs for robustness in identifying these two breeds in particular. Both the

pedigreed American Pit Bulls and Rottweilers formed closed clusters in the STRUCTURE analysis, and, accordingly their breed assignments based on the GeneClass analysis, were 100% successful. This demonstrates that the specific breed categories to which pedigreed Pit Bulls and Rottweilers (as well as the other pedigreed dogs) belong are identifiable based on genetic tests using the 18 STRs reported here. The assignment probability of Pit Bull crosses and Rottweiler crosses were lower, especially for animals assigned to more than two breeds. Furthermore, unlike Himmelberger et al.'s (2008) conclusion, our data support that a mixed breed dog can be assigned with some degree of confidence to its predominant breed/s as reported by its owners.

As illustrated by the complex patterns of the STRUCTURE analysis, the assignments of mixed breed dogs were indistinct, reflecting their mixed ancestry. The mixed breed dogs used in this study represent 43 different breeds and reflect varying degrees of admixture of the breeds. While some of these dogs have been described by their owners and/or have been genetically tested as belonging to a particular breed, they cannot be classified as pedigreed dogs because dog registries require a documented pedigree history and conformation to strict breed standards before recognizing a dog's pedigreed status. Therefore, many of these dogs described as belonging to a specific breed in this study might actually be of mixed ancestry and not breed true-to-type. Also discrepancies in breed definitions could contribute to uncertainty; for example, in the U.S. the American Pit Bull is sometimes called the Staffordshire Bull Terrier. The rich genetic history of the mixed breed dogs is also evident in their heterozygosity values, which are the highest estimates obtained in this study.

Pairwise comparisons of genetic differentiation (Fst) also demonstrate that mixed breed dogs have a greater degree of genetic similarity with all the other breeds and therefore reflect a shared ancestry with them. In order to conclusively characterize the heritage of a dog of unknown breed origin, specifically selected, breed-informative single nucleotide polymorphisms (SNPs) would be more useful than (Parker et al. 2004) these 18 STR loci.

The STR loci studied here appear to be useful in identifying the breed of a particular animal that is the subject of litigation but of greater value is their ability to individualize the DNA of a dog. To evaluate the power of the 18 STRs, a hypothetical canid evidentiary STR profile was used to estimate RMP with Fst correction values of 0.002 and 0.09 for population substructure among the regional populations and among the different breeds, respectively, as recommended by the National Research Council (1996).

Following Budowle et al.'s (2000; 2003) and the National Research Council's (1996) recommendations for source attribution, the probability of not observing the evidentiary profile in a population of N unrelated individuals, or $(1 - RMP)^N$, should be $\geq (1 - \alpha)100\%$ confidence level. Accordingly, for a confidence level of 99% (where α is 0.01) based on the national allele frequencies (from Table 2), RMP values of less than $1 - [(1 - \alpha)1/N] = \alpha/N = 1.429 \times 10^{-10}$ are required to have a high degree of confidence that a profile is unique among the estimated N = 70 million dogs in the U.S. (Budowle et al. 2000; 2003). Estimates of RMP of the fictional profile based on national allele frequencies of 4.8153×10^{-34} with Fst correction of 0.002 for among-region population substructure and of 2.26054×10^{-40} with Fst correction of 0.09 for among-breed population substructure far exceed the threshold for ensuring with a great degree of confidence that the profile is unique among U.S. domestic dogs. Thus, in addition to their utility in identifying the breed composition of an animal that is subject to litigation, the STRs studied here can also identify that animal by its unique genotypic profile with a high level of confidence. Since siblings may share more loci genotypes in common, an appropriate conditional kinship analysis should be performed when necessary. Notably, the proposed reagent kit is also capable of genotyping wolf (*C. lupus*) samples, making it useful in wildlife population and conservation studies as well.

Validation study

While the The Finnzymes Canine 2.1 Multiplex STR Reagent kit holds excellent promise as tool for routine forensic casework involving canine DNA, the results of our study revealed some minor problems that should be brought to the attention of the manufacturer. While the overall PHRs for most markers (excluding the Zinc Finger and FH2017) ranged between 78.57 and 96.14%, intracolor balance was not as robust as in Identifiler® (between 32-45% depending on color channel). Intercolor balance was 22% averaged over 61 samples evaluated. These lower values are due to one or two loci per color yielding too little or too much product. To obtain better intracolor balance, increasing the signal of markers that tended to produce the lowest RFU values per color (FH2017, PEZ05, FH2107, and FH2088) and reducing the signal in markers that tended to generate the highest RFU values (FH2010, Zinc Finger, FH2054) could be investigated. Furthermore, certain primers had high sequence similarity with non-canine species, but this may not be problematic because of the rare incidences that DNAs from these species may co-occur. Also, complete profiles of the non-target species will not be obtained, and non-canid alleles do have distinguishable sizes compared to canine. In particular, the FH2017 forward primer has 95-100% similarity to numerous locations in the horse, human, and rhesus monkey genomes. Also, it is probable that the primers designed for locus FH2017 bind to a region of DNA that contains site variants, which is apparent in the high standard deviation for the peak height ratios calculated. In order to improve performance, a degenerate primer could be added to the reaction. Alternatively, rather than addressing the various problems associated with marker FH2017, the primer pair could be removed from the multiplex altogether. We also discovered that the positive control animal sample provided in the kit is at a lower concentration (0.1 ng/μL) than what is reported in the kit (1 ng/μL). In addition, three peaks were seen in locus FH2328 for the positive control. This peak could potentially be an artifact peak (possibly elevated stutter) generated by the Finnzymes 19-plex, or it could be a tri-allelic pattern associated with the cell line used for positive control F-863. Only the positive control sample exhibited this three peak pattern at locus FH2328. While the original goals of the project have been met, an allelic ladder system would greatly enhance the PCR reagent kit. All commercial human forensic kits include ladders, and if practitioners in the field are to use the canine kit, they must have access to one that is based on standards to which they are accustomed.

Nomenclature

The 18 STR loci that have been assembled into the canine STR reagent kit can be divided into simple, compound, and complex repeat structures. In general, our data agrees with the repeat structure of the defined loci. The exceptions are loci PEZ05, PEZ16, PEZ17, FH2004, and FH2017 where the original sequences were not interpreted according to ISFG convention.

Sometimes mutations in the flanking regions can cause PCR failure and allelic dropout among some of the canine loci studied here (as observed by Dayton [2008]). For some loci, sequence variation in the regions flanking the core repeat region also can impact the PCR product sizes and thus allele designation.

The human STR nomenclature is based on the 1994 recommendations of the ISFG. Standardized canine STRs and validated analysis protocols from their studies as well as the one described herein will promote the use of validated STR panels and robust nomenclatures. The benefits are consistency through time within a laboratory, an ability for inter-laboratory sharing of typing results, pooling of population data, and sharing of experiences to more expeditiously improve processes. The allelic nomenclature proposed here in conjuction with breedwise and regional-wide population genetic study using the markers included in the reagent kit by S. Kanthaswamy and colleagues (submitted) and the developmental validation of the kit (Dayton 2008) will provide a more accurate description of STR allele frequency distributions among the various canine subpopulations, which will enhance the accuracy and precision of canine forensic genetic testing.

This study focused on the development and application of the panel of loci included in the proposed Finnzymes multiplex and the associated genetic database in forensic genetic identity and parentage testing in the U.S. The kit's panel of 18 STRs was shown to be informative and robust for identity testing of canines, and DNA extracted from both blood and buccal cells reliably produced full profiles when amplified with the canine STR kit. Furthermore, the database, which is constructed based on the 18 STRs, is more comprehensive than other dog STR databases in terms of regional representation of pedigreed and mixed breed dog populations in the U.S. The genetic profiles and allele frequencies of important dog breeds in the U.S., which are popular as house pets and could be dangerous as vicious animals linked to fatal dog bites, are also represented in the database.

The Finnzymes multiplex is comparable to other forensic typing systems in its level of sensitivity, ability to detect low level mixture ratios, and power of discrimination while amplifying case-type samples. With enhanced informativity and efficiency, as well as their easy accessibility to the forensic laboratories, the kit and the accompanying population genetic database should combine to form a valuable resource that could potentially develop into a universally accepted canine forensic STR system. With the availability of such a commercial kit, more population data will likely be generated that will enable more precise estimates of the effects of canine population substructure. The data and database presented in these validation and population genetic studies should prove highly useful for the forensic science community as well as for the manufacturer of the kit so that final modifications can be made in order to meet the needs and requirements of the forensic science community.

V. References:

Aasen E, Medrano JF 1990. Amplification of the ZFY and ZFX genes for sex identification in humans, cattle, sheep and goats. Bio/Technology, 8: 1279–1281.

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS, Automated binning of microsatellite alleles: problems and solutions. Molecular Ecology Notes 2007; 7: 10-14.

Bannasch DL, Bannasch MJ, Ryun JR, Famula TR, Pedersen NC 2005. Y chromosome haplotype analysis in purebred dogs. Mammalian Genome, 16; 273-280.

Baute DT, Satkoski JA, Spear TF, Smith DG, Dayton MR, Malladi VS, Goyal V, Kinaga JL, Kanthaswamy S. Analysis of forensic SNPs in the canine mtDNA HVI mutational hotspot region. Journal of Forensic Science, 53:1325-1333. DOI 10.1111/j.1556-4029.2008.00880.x.

Barber MD, McKeown BJ, Parkin BH. Structural variation in the alleles of a short tandem repeat system at the human alpha fibrinogen locus. Int J Legal Med 1996;108:180-185.

Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B. DNA recommendations: Further report of the DNA Commission of the ISFG regarding the use of short tandem repeat systems. Int J Legal Med 1997;110:175-176.

Brauner P, Reshef A, Gorski A. DNA profiling of trace evidence: Mitigating evidence in a dog biting case. J Forensic Sci 2001;46(5):1232-4.

Breen M, Jouquand S, Renier C, Mellersh CS, Hitte C, Holmes NG, Cheron A, Suter N, Vignaux F, Bristow AE, Priat C, McCann E, Andre C, Boundy S, Gitsham P, Thomas R, Bridge

WL, Spriggs HF, Ryder EJ, Curson A, Sampson J, Ostrander EA, Binns MM, Galibert F. Chromosome-Specific Single-Locus FISH Probes Allow Anchorage of an 1800-Marker Integrated Radiation-Hybrid/Linkage Map of the Domestic Dog Genome to All Chromosomes. Genome Research 2001;11:1784-1795.

Brewer D J, Terence C, Phillips A. DOGS IN ANTIQUITY Anubis to Cerberus: The Origins of the Domestic Dog. Aris & Phillips (2002).

Budowle B, Chakraborty R, Carmody G, Monson KL 2000. Source attribution of a forensic DNA profile. Forensic Science Communications. Available at: http://www.fbi.gov/programs/lab/fsc/backissu/july2000/source.htm.

Budowle, B, Hobson, D.L, Smerick, J.B, Smith, J.A.L.: Low Copy Number - Consideration and Caution. In: Twelfth International Symposium on Human Identification 2001, Promega Corporation, Madison, Wisconsin, 2001. Available: http://www.promega.com/ussymp12proc/default.htm.

Budowle B 2003. Population studies on 17 STR loci routinely used in forensic analyses. In: Progress in Forensic Genetics. Brinkmann B, Carracedo A eds.), Elsevier, Amsterdam, pp. 71-74.

Budowle B, Garofano P, Hellmann A, Ketchum M, Kanthaswamy S, Parsons W. van Haeringen W, Fain S, Broad, T 2005. Recommendations for Animal DNA Forensic Identity Testing. International Journal of Legal Medicine. 119: 295 – 302.

Butler J. Forensic DNA typing: biology and technology behind STR markers. (2005) Academic Press, San Diego.

Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA. Developmental Validation of a Single-Tube Amplification of the 13 CODIS STR Loci, D2S1338, D19S433, and Amelogenin: The AmpFiSTR® Identifiler® PCR Amplification Kit. J Forensic Sci 2004; 49(6):1-13.

Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153:1989-2000.

Dayton M. Developmental Validation of an STR Reagent Kit for Forensic DNA Profiling of Canine Biological Material. University of California Master of Forensic Thesis (2008).

DNA Advisory Board (DAB) Recommendations—1994 report concerning further recommendations of the DNA Commission of the ISFG regarding PCR-based polymorphisms in STR (short tandem repeat) systems. Int J Legal Med 1994;107:159-160 and Forensic Sci Int 1994; 69:103-104.

DeNise S, Johnston E, Halverson J, Marshall K, Rosenfeld D, McKenna S, et al. 2004. Power of exclusion for parentage verification and probability of match of identity in American Kennel Club breeds using 17 canine microsatellite markers. Animal Genetics 35:14–7.

Eichmann C, Berger B, Reinhold M, Lutz M, Parson W. Canine-specific STR typing of saliva traces on dog bite wounds. Int J Legal Med 2004a;118:337-342.

Eichmann C, Berger B, Parson W. A proposed nomenclature for 15 canine-specific polymorphic STR loci for forensic purposes. Int J Legal Med 2004b;118:249-266.

Eichmann C, Berger B, Steinlechner M, Parson W. Estimating the probability of identity in a random dog population using 15 highly polymorphic canine STR markers. Forensic Science International 2005; 151(1):37-44.

Evans JJ, Wictum EJ, Penedo MCT, Kanthaswamy, S. Real-Time PCR Quantification of Canine DNA. Journal of Forensic Science 2007; 52: 93 – 96.

Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 2003; 164:1567–1587.

Fisher RA 1932. Statistical methods for research workers. London: Oliver and Boyd 1932.

Francisco LV, Langston AA, Mellersh CS, Neal CL, Ostrander EA 1996. A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. Mammalian Genome 7:359–362.

Gershman KA, Sacks JJ, Wright JC 1994. Which dogs bite? Case-control study of risk factors. Pediatrics 93:913–917.

GeneClass2: A software for genetic assignment and first-generation migrant detection. Journal of Heredity 95:536-539.

Guyon R, Lorentzen TD, Hitte C, Kim L, Cadieu E, Parker HG, Quignon P, Lowe JK, Renier C, Gelfenbeyn B, Vignaux F, DeFrance HB, Gloux S, Mahairas GG, Andre C, Galibert F, Ostrander EA2003. A 1Mb Resolution Radiation Hybrid Map of the Canine Genome. Proceedings of the National Academy of Sciences 100: 5296-5301.

Halverson JA, Basten C. A PCR Multiplex and Database for Forensic DNA Identification of Dogs. J Forensic Sci 2005;50:1-12.

Halverson J, Dvorak J, Stevenson T, inventors. Metamorphix, Inc, assignee. Microsatellite Sequences for Canine Genotyping. US Patent #5,874,217. February 1999.

Hellmann AP, Rohleder U, Eichmann C, Pfeiffer I, Parson W, Schleenbecker U. A Proposal for Standardization in Forensic Canine DNA Typing: Allele Nomenclature of Six Canine-Specific STR Loci. J Forensic Sci 2006;51:274-281.

Himmelberger AL, Spear TF, Satkoski JA, George DA, Garnica WT, Malladi VS, Smith, DG, Allard MW, Kanthaswamy S 2008. Forensic utility of the mitochondrial hypervariable 1 region of domestic dogs, in conjunction with breed and geographic information. Journal of Forensic Science 53:81 – 89.

Hudlow WR, Date Chong M, Swango KL, Timken MD, Buoncristiani MR. 2007. A quadruplex real-time qPCR assay for the simultaneous assessment of total human DNA, human male DNA, DNA degradation and the presence of PCR inhibitors in forensic samples: A diagnostic tool for STR typing. Forensic Science International: Genetics. 2008; 2: 108-125.

Irion DN, Schaffer AL, Famula TR, Eggleston ML, Hughes SS, Pedersen NC 2003. Analysis of genetic variation in 28 dog breed populations with 100 microsatellite markers. Journal of Heredity 94:81–7.

Jombart T 2008. adegenet: an R package for the multivariate analysis of genetic markers. Bioinformatics, submitted.

Kanthaswamy S, Von Dollen A, Kurushima JD, Alminas O, Rogers J, Ferguson B, Lerche NW, Allen PC, Smith DG. (2006). Microsatellite Markers for Standardized Genetic Management of Captive Colonies of Rhesus Macaques (Macaca mulatta). Amer J Primatol. 68:73-95.

Kanthaswamy S, Tom B.K, Mattila A-M, Johnston E, Dayton M, Kinaga J, Erickson B, J-A, Halverson J, Fantin D, DeNise S, Kou A, Malladi V, Satkoski J, Budowle B, Smith DG, Koskinen MT. (submitted). Canine Population Data Generated from a Multi-Plex STR Kit for Use in Forensic Casework. Journal of Forensic Science.

Klein SB, Wallin JM, Buoncristiani MR. Addressing Ambient Temperature Variation Effects on Sizing Precision of AmpFISTR® Profiler Plus™ Alleles Detected on the ABI Prism® 310 Genetic Analyzer. Forensic Sci Comm.2003; 5(1).

Koskinen MT, Bredbacka P. A convenient and efficient microsatellite-based assay for resolving parentages in dogs. Anim Genet. 1999; 30(2):148-9.

Kou A, Kanthaswam S (n prep.) Generation Canine Parentage Analysis Using Microsatellite Markers. Animal Genetics.

Leibelt C, Budowle B, Collins P, Daoudi Y, Moretti T, Nunn G, Reeder D, Roby R. Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles. Forensic Science International 2003; 133: 220-227.

Lu DJ, Sun HY, Chen LX. Genotype reliability of short tandem repeat typing from minute DNA. Fa Yi Za Zhi 2003;19(3):151-3.

Meiboom M, Murua Escobar H, Winkler S, Nolte I, Bullerdiek J. Molecular characterization and mapping of the canine KRAB zinc finger gene ZNF331. Anim Genet 2004; 35(3):262-3.

Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K, Wiegand NA, Francisco LV, Gibbs M, Aguirre GD, and Ostrander EA. A linkage map of the canine genome. Genomics 1997; 46:326-36.

Mellersh CS, Hitte C, Richman M, Vignaux F, Priat C, et al. 2000. An integrated linkage-radiation hybrid map of the canine genome. Mammalian Genome 11:120–30.

Menotti-Raymond, M, Davis, V, O'Brian, S 1997. Pet cat hair implicates murder suspect. Nature 376: 774.

Menotti-Raymond M, David V, Wachter L, Butler J, O'Brien S. An STR Forensic Typing System for Genetic Individualization of Domestic Cat (Felis catus) Samples. J Forensic Sci Sept. 2005; 50(5): 1061-1070.

Neff MW, Broman KW, Mellersh CS, Ray K, Acland GM, Aguirre GD, Ziegle JS, Ostrander EA, and Rine J. A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. Genetics 1999;151:803-820.

Ostrander EA 2003. A 1Mb Resolution Radiation Hybrid Map of the Canine Genome. Proceedings of the National Academy of Sciences 100: 5296-5301.

Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD,Malek TB, Johnson GS, DeFrance HB, Ostrander EA, Kruglyak L 2004**.** Genetic structure of the purebred domestic dog. Science 304:1160-1164.

Pádár Z, Angyal M, Egyed B, Füredi S, Woller J, Zöldág L, Fekete S. Canine microsatellite polymorphisms as the resolution of an illegal animal death case in a Hungarian Zoological Gardens. Int J Legal Med 2001a;115:79-81.

Pádár Z, Egyed B, Kontadakis K, Zoldag L, Fekete S 2001b. Resolution of parentage in dogs by examination of microsatellites after death of putative sire: case report. Acta Vet Hung 49:269–273.

Pádár Z, Egyed B, Kontadakis K, Füredi S, Woller J, Zöldág L, Fekete S. Canine STR analyses in forensic practice: Observation of a possible mutation in a dog hair. Int J Legal Med 2002;116:286-288.

Phusion Hot Start High-Fidelity DNA Polymerase, New England Biolabs, Finnzymes, Version 1.2 September 2006.

Piry S, Alapetite A, Cornuet, J-M, Paetkau D, Baudouin L, Estoup A. (2004)

Primmer CR, Møller AP, Ellegren H. A wide-range survey of cross-species microsatellite amplification in birds. Mol Ecol. 1997; 6(1):101)

Pritchard JK, Stephens M, Donnelly P 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Raymond M, Rousset F 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. Heredity, 86: 248 – 249.

Rosenblum BB, Oaks F, Menchen S, Johnson B. Improved single-strand DNA sizing accuracy in capillary electrophoresis. Nucleic Acids Research 1997; 25(19):3925-3929.

Rice WR 1989. Analysis tables of statistical tests. Evolution 43: 223-225.

Sacks JJ, Sinclair, L Gilchrist J, Golab GC, Lockwood, R 2000. Breeds of dogs involved in fatal human attacks in the United States between 1979 and 1998. JAVMA 217 (6): 836-840.

Savolainen P, Zhang Y, Luo J, Lundeberg J, Leitner T. Genetic Evidence for an East Asian Origin of Domestic Dogs 2002. Science; 298:1610-1613.

Shibuya H, Collins BK, Huang TH, Johnson GS 1994. A polymorphic (AGGAAT)n tandem repeat in an intron of canine von Willebrand factor gene. Animal Genetics 25:122.

Shutler GG, Gagnon P, Verret G, Kalyn H, Korkosh S, Johnston E, Halverson J. 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, 1999; 44(3): 623-626.

Smalling BB, Satkoski JA, Tom BK, Wing YS, Erickson BJ-A, Spear TF, Smith DG,,Kanthaswamy S. (in prep). The significance of regional and mixed breed canine mtDNA databases in forensic science.

Tom BK, Koskinen MT, Dayton MR, Mattila A-M, Johnston E, Fantin D, DeNise S, Spear T, Smith DG, Satkoski J, Budowle B, Kanthaswamy S (submitted). Development of a Nomenclature System for a Canine STR Multiplex Reagent Kit. International Journal of Legal Medicine.

Weir BS, Cockerham CC 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

National Research Council 1996. The evaluation of forensic DNA evidence. Washington, DC: National Academy Press.

Nord K, Gunneriusson E, Ringdahl J, Stahl S, Uhlen M, Nygren P. Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. Nature Biotechnol. 1997; 15, 772-777.

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.

U.S. Pet Ownership and Demographics Sourcebook (2007 Edition), Schaumburg, Ill.: Center for Information Management, American Veterinary Medical Association, ©1997.

Wang Y, Prosen DE, Mei L, Sullivan JC, Finney M, Vander Horn PB. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res. 2004 32(3):1197-207.

Wetton JH, Higgs JE, Spriggs AC, Roney CA, Tsang CSF, Foster AP. Mitochondrial profiling of dog hairs. Forensic Science International 2003;133:235-241.

Wikman M, Steffen A, Gunneriusson E, Tolmachev V, Adams GP, Carlsson J, Stahl S. Selection and characterization of HER2/neu-binding affibody ligands. Protein Eng, Des. Sel. 2004; 17(5), 455-462.

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.

VI. Dissemination of Research Findings:

The underlying scientific approaches employed in the validation procedure, findings from the canine population genetic research [including discriminating statistics, demographic data and analysis methods], allele sequences, reaction composition, and thermal cycling parameters have been made public through publication in peer-reviewed forensic literature such as the Journal of Forensic Science and the International Journal of Legal Medicine. We have also disseminated the procedures and results of our population and developmental studies to local and international state and federal academic and forensic communities through professional meetings.

Publications:

Tom, BK, Koskinen, MT, Dayton, MR, Mattila, A-M, Johnston, E, Fantin, D, DeNise, S, Spear, T, Smith, DG, Satkoski, J, Budowle, B, Kanthaswamy, S (submitted). Development of a Nomenclature System for a Canine STR Multiplex Reagent Kit. *International Journal of Legal Medicine*.

Dayton M. Developmental Validation of an STR Reagent Kit for Forensic DNA Profiling of Canine Biological Material. University of California Master of Forensic Thesis (2008).

Kanthaswamy, S, Tom, BK, Mattila, A-M, Johnston, E, Dayton, M, Kinaga, J, Erickson, B, J-A, Halverson, J, Fantin, D, DeNise, S, Kou, A, Malladi, V, Satkoski, J, Budowle, B, Smith, DG, Koskinen, MT (submitted). Canine Population Data Generated from a Multi-Plex STR Kit for Use in Forensic Casework. *Journal of Forensic Science.*

Baute DT, Satkoski JA, Spear TF, Smith DG, Dayton MR, Malladi VS, Goyal V, Kinaga JL, Kanthaswamy S. Analysis of forensic SNPs in the canine mtDNA HVI mutational hotspot region. Journal of Forensic Science 2008; DOI 10.1111/j.1556-4029.2008.00880.x.

Himmelberger, AL, Spear, T F, Satkoski, J A, George, D A, Garnica, W T, Malladi, VS, Smith, DG, Allard MW and Kanthaswamy, S. Forensic utility of the mitochondrial hypervariable 1 region of domestic dogs, in conjunction with breed and geographic information (2008). *Journal of Forensic Science*, 53:81 – 89.

Evans JJ, Wictum EJ, Penedo, MCT and Kanthaswamy S (2007). Real-Time PCR Quantification of Canine DNA. *Journal of Forensic Science*, 52: 93 – 96.

Oral and poster presentations:

Kanthaswamy, S, Bradley Tom, Melody Dayton, Briana Smalling. *The Development and Validation of a Standardized Canine STR Panel for Use in Forensic Casework*. Poster presentation at the XXXI Conference of the International Society for Animal Genetics, RAI Conference Center, Amsterdam, The Netherlands. July 20 – 24, 2008.

Bradley Tom, Melody Dayton, Briana Smalling, Kanthaswamy, S. *The Development and Validation of a Standardized Canine STR Panel for Use in Forensic Casework*. Poster presentation at the 9th Annual US National Institute of Justice (NIJ) Conference 2008, Arlington, Virginia, July 20 – 24, 2008.

Kanthaswamy, S, Bradley Tom, Melody Dayton, Briana Smalling. *The Development and Validation of a Standardized Canine STR Panel for Use in Forensic Casework*. Poster

presentation at the XXXI Conference of the International Society for Animal Genetics, RAI Conference Center, Amsterdam, The Netherlands. July 20 – 24, 2008.

Kanthaswamy, S. *Progress report on the Development and Validation of a Standardized Canine STR Panel for Use in Forensic Casework.* US National Institute of Justice 7th Annual Grantees Workshop, Arlington, Virginia. June 26 – 28, 2006.

Kanthaswamy, S. *Animal Forensics – Use of mtDNA in Forensics Casework.* Mitochondrial DNA Analysis and Typing (R602). Jan Bashinski DNA Laboratory, Department of Justice, State of California, Richmond, CA. February 9, 2006.

Kanthaswamy, S. *Progress report on the Development and Validation of a Standardized Canine STR Panel for Use in Forensic Casework.* US National Institute of Justice 6th Annual Grantees Workshop, Washington DC. June 26 – 29, 2005.

Kanthaswamy, S. *Veterinary Forensics – Canine Forensics Evidence.*
American Kennel Club (AKC)/Canine Health Foundation/UC Davis Center for Comparative Animal Health (CCAH), University of California, Davis, CA. May 21, 2005.

Kanthaswamy, S. *CSI Davis - Using Animal DNA Evidence to Fight Crime.*
Sacramento City College, Sacramento, CA. April 29, 2005.

Kanthaswamy, S. *Veterinary forensics – a tool for crime fighters.* California Crime Laboratory Directors Meeting, Bahai Hotel, San Diego, CA. April 7, 2005.

Kanthaswamy, S. *Animal CSI - Using Animal DNA Evidence to Fight Crime.*
UC Davis Division of Biological Sciences Biology Undergraduate Scholars Program – Honors Research (BUSP-HR). March 9, 2005.

Kanthaswamy, S. *Veterinary Forensics in 3D - Establishing A Veterinary Forensics Program at the UC Davis School of Veterinary Medicine: A Job Talk*. Department of Medicine and Epidemiology, School of Veterinary Medicine, UC Davis. December 12, 2004.

Kanthaswamy, S. *Veterinary Forensics – Use of Animal DNA in Forensics Casework*. Jan Bashinski DNA Laboratory, Department of Justice, State of California, Richmond, CA. November 11, 2004.

Kanthaswamy, S. *Animal CSI - Using Animal DNA Evidence to Fight Crime.* General Education Class (GED), Charles A. Jones Skill and Business Education Center, Sacramento City Unified School District, Sacramento, California, October 20, 2004.

Kanthaswamy, S. *Introduction to Animal Forensics.* Roundtable Forum on Animal Forensics at the International Society for Animal Genetics Conference, Tokyo, Japan, September 11 to September 16, 2004.

Kanthaswamy, S, J. Evans and E. Wictum. *Improved Genotyping Through the Use of Real-Time PCR DNA Quantification.* Poster presentation at the International Society for Animal Genetics Conference, Tokyo, Japan, September 11 to September 16, 2004.

Acknowledgements:

Many have contributed to this work in terms of sampling, research in the laboratory and the write up of research articles that have led to this final report. Some of the notable contributors are:
Mikko Koskinen, David Glenn Smith, Bruce Budowle, Joy Halverson, Melody Dayton, Anna-Maria Matilla, Brad Tom, Briana Smalling, Jennifer Kinaga, Jessica Satkoski, Terry Spear, Eric Johnston, Sue deNise, Dennis Fantin, Walther Parsons, Cordula Eichmann, Andreas Hellmann, Shelia Stevens, Shannon Flynn, and John Butler. Without their contribution, this work would not have happened.

We wish to thank Dr Lois Tully for her steadfast support for this project.