Forensic Use of Short Tandem Repeats *via* PCR

C.T. Caskey and H.A. Hammond

Institute for Molecular Genetics and Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX 77030

INTRODUCTION

The DNA sequence of man's chromosomes has the potential for wide variation from individual to individual since only 1% encodes functional elements, *i.e.* genes. Sequence differences were first detected by restriction fragment length polymorphisms (RFLPs) using the Southern method (Southern, 1975). These two-allele polymorphisms were used for mapping Huntington disease locus and diagnosis of both ß-thalassemia and sickle cell disease. Jeffreys (1985b) and Nakamura (1987) subsequently described a new class of RFLPs which arise from repeated sequences occurring in different copy numbers from individual to individual. They were referred to as satellite sequences or variable number of tandem repeats (VNTR). These highly informative repeat sequence polymorphisms have been used extensively in the mapping of human disease genes such as cystic fibrosis (CF) and neurofibromatosis (NF), and more relevant to this meeting, for personal identification (Jeffreys *et al.* 1985a) (Gill *et al.* 1985). The method of identification was Southern analysis. At many of these loci, the repeat sequences are smaller than the resolving power of agarose gels, so discrete alleles will not be differentiated. Despite this shortcoming, the large number of alleles provides a very powerful DNA-based personal identification method. A summary of how such alleles are measured and grouped for the purpose of determining RFLP matching and significance of match has been discussed in numerous articles (Budowle, 1990; Budowle *et al.* 1991; Devlin *et al.* 1991). Occasionally a forensic DNA specimen may migrate in the gel faster or slower than expected, thus assigning an allele to an incorrect molecular weight and possibly leading to incorrect matching or mismatching. While it has been suggested that internal molecular weight markers could allow for correction of "band shifts", a general method for this correction is not validated. The uncontrollable character of the forensic specimen has also limited the VNTR/Southern technology, since high molecular weight DNA is required in substantial quantity (Budowle *et al.* 1990; McNally *et al.* 1989; Baechtel, 1988). Consequently as high as 25% of forensic specimens are useless for analysis.

We observed in our sequencing of the HPRT gene a single tetramer repeat (AGAT) which was found to vary from person

Advances in Forensic Haemogenetics 4
Edited by Ch. Rittner and P. M. Schneider

to person in the number of tetramer repeats (Edwards *et al.* 1990). From this initial observation we have proceeded to develop and characterize a series of highly polymorphic loci whose polymorphisms occur on the basis of single tri-, tetra-, and pentanucleotide repeats. We believe these short tandem repeats (STRs) to be extremely useful genetic markers for personal identification and have shown them to be useful for forensic analysis.

ACQUISITION OF LOCI

The genetic loci listed in Table 1 are now under study since we know them to be highly polymorphic. For some, the repeated sequence was discovered by DNA sequencing of genes being studied in our laboratory. For others, the sequences were identified by scanning Genbank entries. Finally, we have now developed methods for specifically seeking out such polymorphic loci. There appears to be no limitation on the identification of many polymorphic loci.

Table 1. Current STR loci for forensic use

Locus and STR	Chromosome
HUMRENA4[ACAG]	1q32
HUMFABP[AAT]	4q31
HUMTH01[AATG]	11p15.5
HUMPLA2A1[AAT]	12q
HUMCD4[AAAAT]	12p
HUMHPRTB[AGAT]	Xq26
HUMARA[AGC]	Xcen-q13

Two methods are now being used for identification of new STR loci. Since we do not seek a specific chromosome localization of an STR and we are early in development of markers, a shotgun approach continues to be the most rapid. Sheared human DNA is cloned into a DNA sequencing vector, plated at low density and those containing one of five repeats identified by hybridization of a radioactive oligomer of 30 bases in length. Using stringent conditions only those predicted to have more than eight repeats in tandem are detected. Sequencing the ends of the clone provides the flanking sequence to the STR, necessary for fashioning a pair of primers for polymerase chain reaction (PCR) amplification. While the method is useful, it has the disadvantage that the sequence information may be inadequate to develop small PCR

elements, *i.e.* the primers may be distant to the repeat. For this reason we developed a general method which provides sequence data immediately flanking the STR (Edwards *et al.* 1991a). We have used the method to obtain 16 new STR loci. Each was identified in λ clones. Our success at amplifying both sides of the AAT repeat is greater than 70%. We have obtained sequence information for 8 pairs of PCR primers and found the repeat sequence to amplify in all cases. Presently 1 of 6 is polymorphic as determined by survey of a panel of 10 unrelated female individuals. Examples of a highly polymorphic and a monomorphic STR are given in Fig. 1.

POPULATION GENETIC FEATURES

The establishment of reliable databases for STRs is operationally easier than for VNTR loci, since the method is PCR-based, alleles are distinguished by sequencing gel analysis, and the allele number is generally smaller. An example of four databases for the AGC repeat within the coding of the androgen receptor is given in Fig. 2. In a separate communication we reported the population genetic features of five loci. There have been no new mutations observed for these loci when over 900 meioses were studied in families provided by Centre d'Etude du Polymorphisme Humain (CEPH). Thus the inheritance of our STRs is stable and they are suitable for parentage studies. We have had the experience of observing new mutation events in CA repeats at the Duchenne muscular dystrophy locus (DMD) and therefore have reservations regarding their stability and utility for parentage studies. Four ethnic population databases have been examined for HUMFABP[AAT], HUMHPRTB[AGAT], HUMARA[AGC], HUMTH01[AATG] and HUMRENA4[ACAG] to determine their population genetic features. The details of these studies are given elsewhere (Edwards *et al.* 1991b). We conclude, however, that the frequencies of alleles measured reflect a genetic marker system which is conforming to Hardy-Weinberg equilibrium and thus allele frequency measurements can be used to calculate the significance of allele matches between two DNA specimens. Furthermore, we find no evidence of linkage disequilibrium between loci and thus the value at each loci identification can be multiplied to estimate the significance of match. These features indicate that these loci can be used in forensic studies based on their genetic properties.

MULTIPLEX AMPLIFICATION OF STRs

We have shown that multiplex amplification of as many as 9 target loci (exons) within the DMD gene can be achieved with fidelity (Chamberlain *et al.* 1988). The multiplex amplification of three STRs is readily achieved using radioisotopic

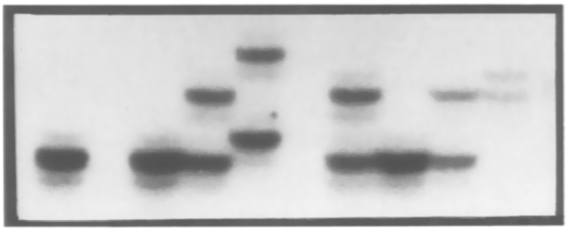


Fig. 1A. PCR results of a polymorphic locus with the STR repeat AAT for ten unrelated female individuals

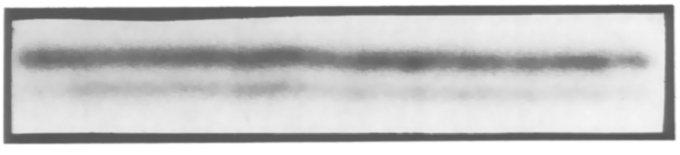


Fig. 1B. PCR results of a monomorphic locus with the STR repeat AAT for ten unrelated female individuals

detection. We sought to co-amplify 9-12 loci and thus turned to a fluorescent detection system. This analysis was facilitated by use of the ABI 373 automated DNA sequencer which uses sequencing gels for resolution and provides 4 different fluorescent tags within the same lane. By judicious choice of fluorescent labelled primers, knowledge of allele sizes, and primer spacing we have achieved a 3 locus co-amplification and related the products to internal allele standards (Edwards *et al*. 1991a). More recently, ABI has developed a GeneScanner which appears to have advantages for this type of analysis. This device uses a molecular weight ladder from prokaryotic sources which allows allele identifications and molecular weight determinations. This appears to be the proper direction to move, toward a user friendly, automated analytic instrument. The A.L.F. unit manufactured by Pharmacia/LKB also has features making it amenable to gel analysis. We are adapting our most informative loci to a multiplex PCR amplification procedure with fluorescent products.

FORENSIC APPLICATIONS

We have applied our STR personal identification system to several medical diagnostic and forensic circumstances. We

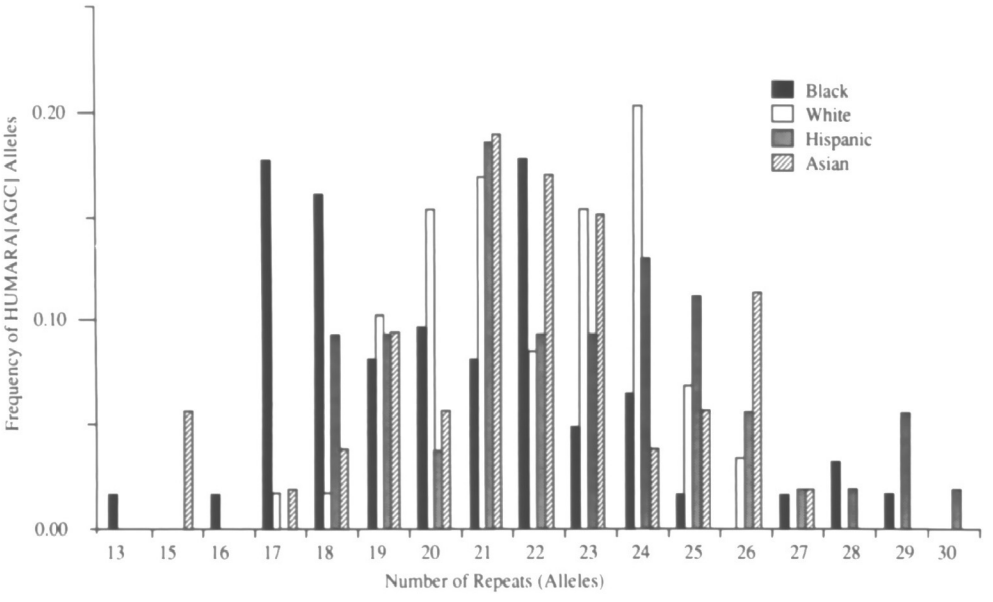


Fig. 2. Allele frequency graph of caucasian, black, hispanic, and asian populations for the HUMARA[AGC] locus

have been able to provide quick resolution of questions in zygosity testing, maternal or fetal origin of chorionic villus biopsies, and pedigree validation for linkage analysis. It has been surprising how quickly we have received a number of request for prenatal paternity testing. Our program has thus far only provided prenatal paternity testing for those case related to documented sexual assaults. We have completed four such cases to date. Counseling in all cases revealed the desire to continue the pregnancies in the event that the fetus was their husband's. One case, shown in Fig. 3, indicated that the fetus of a victim of gang rape was not the progeny of the husband, and termination of the pregnancy was elected. In another case counseling as to the power of the method resulted in the victim dropping the rape charges and identifying a consensual partner.

Forensic identification services have been provide to the Armed Forces Institute of Pathology by a Collaborative effort with Cellmark Diagnostics. The circumstances of aircraft crew submersion (1 month), scud and weapon explosions and the return of air crew remains allowed for the use of PCR based testing to help verify identifications. Fig. 4 shows an

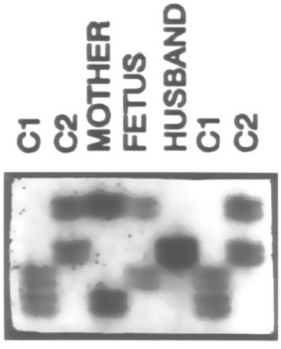
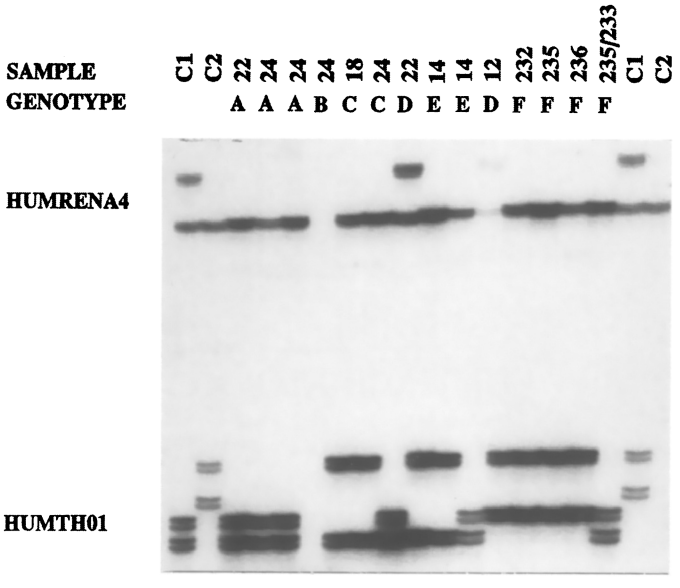


Fig. 3. Results from one of six STR loci analyzed for a paternity trio. The mother was the victim of a gang rape. The husband was excluded as being the father of the fetus using STRs. This exclusion information was the basis for terminating the pregnancy

example of studies performed to verify assignment of body parts by forensic pathologists. Two of the six genetic markers used are illustrated. Note that we observed three genotypes within samples that had been associated with one victim. Calculation of the frequency of six genotypes observed ranged from 1 in 12,000 to 1 in 450,000.
The United States had approximately 450,000 troops involved in the Desert Storm operation, however, all samples were from small combat groups. Had reference samples been available identification would have been greatly simplified. Paternity testing was also employed to verify the remains of a pilot who had been classifies as Missing in Action (MIA). Parental blood specimens were associated with the genotype for the mia remains verifying that the correct body had been returned.

Our laboratory currently provides RFLP-VNTR results to the U.S. Court system, and has reported a significant number (42%) of exclusions in criminal sexual assault cases. We have initiated a policy of screen cases with STRs for rapid identification of exclusions, although are not yet introducing STR data into the courts. All cases and particularly matches with STRs will continue to be studied in parallel with VNTR loci. This experience will strengthen our objective of replacing Southern-based technology, with PCR based analysis of STR loci.



GENOTYPE FREQUENCIES FOR SIX LOCI (CAUCASIAN)

A - 1 IN 450000 B - 1 IN 126500 C - 1 IN 12000
D - 1 IN 83000 E - 1 IN 357000 F - 1 IN 62500

Fig. 4. Results for 2 STR loci on samples from Operation Desert Storm. Numbers above the samples indicate the assignment of samples by the Army. Letters represent the genotypes represented by STR analysis for six loci for each sample

REFERENCES

Budowle B, Baechtel FS, and Adams DE (1990) Validation with regard to environmental insults of the RFLP procedure for forensic purposes. American Chemical Society Symposium Series (in press)

Budowle B, Giusti AM, Waye JS, Baechtel FS, Fourney RM, Adams DE, Presley LA, Deadman HA, and Monson KL (1991) Fixed BIN analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci for use in forensic comparisons. Am J Hum Genet 48:841-855

Budowle BF (1990) Data for forensic matching criteria for VNTR profiles. In: Data acquisition and statistical analysis for DNA typing laboratories. Promega Corporation, Madison, WI, pp. 103-115

Baechtel FS (1988) Recovery of DNA from human biological specimens. Crime Laboratory Digest 15:95-96

Chamberlain JS, Gibbs RA, Ranier JE, Nguyen P-N, and Caskey CT (1988) Deletion screening of the Duchenne muscular dystrophy locus *via* multiplex DNA amplification. Nucleic Acids Res 16:11141-11156

Devlin B, Risch N, and Roeder K (1991) Estimation of allele frequencies for VNTR loci. Am J Hum Genet 48:662-676

Edwards A, Voss H, Rice P, Civitello A, Stegemann J, Schwager C, Zimmermann J, Erfle H, Caskey CT, and Ansorge W (1990) Automated DNA sequencing of the human HPRT locus. Genomics 6:593-608

Edwards A, Civitello A, Hammond HA, and Caskey CT (1991a) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet, in press

Edwards A, Hammond HA, Caskey CT, and Chakraborty R (1991b) Population genetics of trimeric and tetrameric tandem repeats in four human ethnic groups. Genomics, submitted

Gill P, Jeffreys AJ, and Werrett DJ (1985) Forensic application of DNA "fingerprints". Nature 318:577-579

Jeffreys AJ, Wilson V, and Thein SL (1985a) Individual-specific "fingerprints" of human DNA. Nature 316:76-79

Jeffreys AJ, Wilson V, and Thein SL (1985b) Hypervariable 'minisatellite' regions in human DNA. Nature 314:67-73

McNally L, Shaler, RC, Baird M, Balazs I, De Forest P, and Kobilinsky L (1989) Evaluation of deoxyribonucleic acid (DNA) isolated from human bloodstains exposed to ultraviolet light, heat, humidity, and soil contamination. J Forensic Sciences 34:1059-1069

Nakamura Y, Leppert M, O'Connel P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, and White R (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616-1622

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503-517