Evaluation of new Mini STR markers for DNA based human identification: An *in silico* approach

N.D.S Goonawardhana[1], Preethi V. Udagama[2], N.D Fernandopulle[1], Ruwan J. Illeperuma[1]

[1]*Forensic Molecular Unit, Borella*

[2]*Department of Zoology, Faculty of Science, University of Colombo*

Introduction

Analysis of Short Tandem Repeat (STR) markers on the human genome is an important tool in the identification of individuals based on their genetic makeup. DNA extracted from materials of biological evidence is subjected to Polymerase Chain Reaction (PCR) to generate DNA profiles, unique to an individual. However, forensic analysts are challenged to obtain a complete DNA profile by the failure in PCR amplification of such biological evidence due to several factors: for example, high humidity and temperature in the tropics (as in Sri Lanka), promote the degradation of such biological materials and rapidly reduce the possibility of typing such nuclear DNA. Conventional PCR technology used in Sri Lanka analyses amplicons generated in the size range of 150 to 450 nucleotide bases. Hence, DNA in a biological sample fragmented into smaller sizes cannot be amplified by PCR. Therefore, we have evaluated the application of three new miniaturized –Short Tandem Repeat markers (mini-STRs) screened for possible application in degraded DNA evidence analysis. These novel STR markers are capable of generating smaller sized PCR amplicons (less than 150 bp) by PCR. No previous studies have been recorded in the application of these three STR loci.

Materials and Methods

Selection of STR loci

The three Mini-STR markers were selected by screening the STR marker sets published in Marshfield Clinic Center for Medical Genetics (http://research.marshfieldclinic.org/genetics/). The selected STRs, namely D4S2632, D6S2436, and D19S589 have not been adapted to human identity testing up to date. The sequences of the STR loci screened were obtained from a BLAST –nucleotide search on www.ncbi.com. The chromosomal positions were determined by BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) and assembly of the human genome version of Feb2009.

Primer designing

The manual primer designing tool of the standalone "Primer premier 3" software was used to effectively locate primers adjacent to the repeat region of each STR locus. The G and C content of the primer were maintained within the range of 40 % to 60 % and the melting temperatures from 53 ºC -55 ºC. Successfully designing primer sets were evaluated for probable inter-primer interactions using the options in "Primer premier 3" software.

Evaluation of nucleosome forming potentials (NFPs)

Nucleosomes are the basic repeating structural and functional unit of chromatin consisting of nine histone proteins and about 166 bp of DNA (Kogan et al, 2006; Holde, 1988). The algorithm of Nucleosome eXclusion Sensor (NXSensorversion 1.3.1) (http://www.sfu.ca/_ibajic/NXSensor/) was used to determine the nucleosome-free regions of the selected STR marker sequences. The algorithm detects an input sequence for three known nucleosome-free sequences: 10bases of poly-A, 10 bases of poly-T, and a combination of Gs and Cs (A_10, T_10, or [(G/C) 3N2] _3) (Thanakiatkrai et al, 2011). If any of these sequences are detected in the given sequence, the program outputs the sequence in FASTA format and highlights the nucleosome-free region.

The FASTA sequences (~ 100bp up and downstream to the repeat motif) of novel STR markers D4S2632, D6S2436, and D19S589 were entered in to the web based NXSensorversion 1.3.1 and the accessible scores for each STR marker were calculated according to the following formula (Luykx et al, 2006):

$$\text{Accessibility}(OSmin) = \frac{\text{Total length of open contiguous segments } \geq OSmin}{(\text{length of input sequence}) - (\text{total length of ambiguous segments})}$$

Results and Discussion

Screening Candidate Loci

STR Locus	**Chromosomal location**	**Chromosomal pb Position**	**Repeat Motif**	**GenBank accession**	**GenBank allele**	**Mini-STR amplicon length (bp)**	**Accessibility (*OSmin*)**
D4S2632	4p15-p14	35704165	(GATA)n	G08391.1	13	105	0
D6S2436	6q24.1	154136091	(GATA)n	G27284.1	9	91	0
D19S589	19q13.42	58498394	(GATA)n	G08026.1	13	98	0

Table 1—*Information on three novel STRs evaluated in this study. Chromosomal location and base pair (bp) position of each marker was determined by using BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) and the Feb.2009 assembly of the human genome. The miniSTR amplicon length is based on the GeneBank allele observed. The accessibility(OSmin) values for the three STR loci D4S2632, D6S2436 and D19S589*

The potential mini STR markers were screened from previously well characterized STR loci (Ghebranious et al, 2003) including the whole genome screening sets of STRs reported extensively in genetic linkage studies across the entire nuclear genome to determine specific genes that cause or has a linked to human diseases. A subset of the screening markers were also used to characterize the genetic diversity in global populations (Rosenberg et al, 2002 & 2003; Ayub et al, 2003)

When screening the probable STR loci to be used as mini STR markers, we considered several characteristics of each STR locus in order to obtain the best possible loci. Firstly, the STR loci consisting of tetrameric repeat motif were selected, as is strongly recommended in forensic DNA analysis, since these markers are highly polymorphic, with heterozygosity >0.90 and results reduced stutter effect in the PCR reaction (Kimpton et al, 1993). Secondly, the STR loci containing tetrameric repeat motifs were evaluated for "clean flanking regions" upstream and downstream to the repeat motif (approximately

75bp). This approach is very important when the PCR primers are located closer to the repetitive region since a clear flanking region can give rise to good primer hybridization in the PCR reaction. Thirdly, the STR loci that can produce PCR fragments less than 150 bp were selected. Finally, the selected three novel STR loci D4S2632, D6S2436 and D19S589 were evaluated for the nucleosome forming potential (NFPs) by calculating the accessibility (*OSmin*) for each STR marker with the formula (Butler et al, 2003); 1). The accessibility (*OSmin*) values for the STR markers D4S2632, D6S2436 and D19S589 were found to be zero which demonstrates the likelihood of a DNA sequence to be bound by nucleosomes.

Conclusion

This approach outlines the initial efforts to develop mini STR markers that can be used to supplement the loci included in CODIS (Combined DNA Index System) core loci to increase the power of discrimination in analyzing highly degraded DNA evidence in forensic casework. Apart from using these markers for forensic purposes, they can be successfully incorporated in cases where additional markers are needed to conclude the analysis such as human parentage testing or the analysis between closely related individuals.

References:

Ayub, Q et al. 2003. *Am J Phys Anthropol*, 122(3):259–68.

Butler, JM et al 2003. *Forensic Science International*, Vol. 48, No. 5.

Ghebranious N et al 2003. *BMC Genomics*;4(6):1–10.

Kimpton CP et al 1993. *PCR Methods Applications.*, **3**, 13–22.

Kogan S et al 2006. *J. Biomol. Struct. Dyn*, 24 (2006) 43–48.

Luykx P et al Nucleic Acids Research, vol. 34, Web Server issue, pp. W560-W565, July 2006. PMID: 16845070 .

Rosenberg NA et al 2002. *Science* ;298(5602):2381–5.

Rosenberg N et al 2003. *Am J Hum Genetics*,73(6):1402–22.

Thanakiatkrai P et al 2011. Forensic Science International: Genetics 5: 285–290.

Van Holde, K E 1988. Chromatin: *Springer Series in Molecular Biology* (New York,Springer-Verlag,).