Developmental Validation of a Novel 13-loci STR Multiplex Method for *Cannabis sativa* DNA Profiling

Rachel Houston[1] BS*, Matthew Birck[2] PhD, Sheree Hughes-Stamm[1] PhD, David Gangitano[1] PhD

[1]Department of Forensic Science, Sam Houston State University, 1003 Bowers Blvd., Huntsville, TX 77340
[2]New York Laboratory, U.S. Customs and Border Protection, U.S. Department of Homeland Security, 1100 Raymond Blvd, Newark, NJ 07102, USA

U.S. Customs and Border Protection

INTRODUCTION

Forensic DNA typing is typically performed on human DNA samples. However, the molecular analysis of plant DNA is increasingly being studied [1] and considered for use in criminal justice systems around the world [2, 3]. Plant DNA can be used to link a suspect or a victim to a location (crime scene) or in the case of marijuana, can be used to aid in the investigation of drug cases. The development of a validated method using molecular techniques such as short tandem repeats (STRs) for the genetic identification of *C. sativa* will aid in the individualization and origin determination of *Cannabis* samples.

In order to develop a reliable STR method for *Cannabis* identification, the best markers currently available were chosen as a measure of continuity within the field. In choosing markers, dinucleotide repeat markers were avoided. All markers chosen have been previously described using IUPAC nomenclature [4, 5]. Based upon our previous research [10], we improved upon a STR multiplex method by both discarding STR loci with poor performance as well as incorporating six new tetranucleotide markers recently described by Valverde *et. al* [4].

This study describes the developmental validation of a *C. sativa* STR multiplex following guidelines established by the Scientific Working Group on DNA Analysis Methods (SWGDAM). For this purpose, a 13-loci STR multiplex method was developed, optimized, and validated according to relevant ISFG and SWGDAM guidelines.

MATERIALS AND METHODS

DNA Collection DNA from marijuana samples (N=101) was extracted from three previously processed case sets at the U.S. Customs and Border Protection LSS Southwest Regional Science Center. DNA was previously extracted and quantified according to Houston *et. al.* [6].

STR Analysis *Cannabis* STR profiling was conducted in a 13-loci multiplex format modified from a previous study [6]. The multiplex consisted of previously published *Cannabis* STRs including seven markers from a previous multiplex (D02 CANN1, C11 CANN1, B05 CANN1, H06 CANN2, ANUCS305, ANUCS501, and CS1) [6] and six newly proposed tetranucleotide markers (1528, 4910, 5159, 9043, 3735, and 9269) [4]. PCR amplification was performed using the Type-IT Microsatellite PCR Kit (Qiagen) on the Eppendorf Master Cycler Gradient (Eppendorf). Separation and detection of PCR products was performed on the 3500 Genetic Analyzer (Applied Biosystems). Genotyping was performed using a customized bin/panel on the GeneMapper v.5 software (Applied Biosystems).

Allelic Ladder Design Forty *C. sativa* samples were screened initially to determine the variability of alleles observed in the population. Using the most common alleles observed, an allelic ladder was generated according to previous reports [6, 7].

RESULTS & DISCUSSION

Multiplex Design

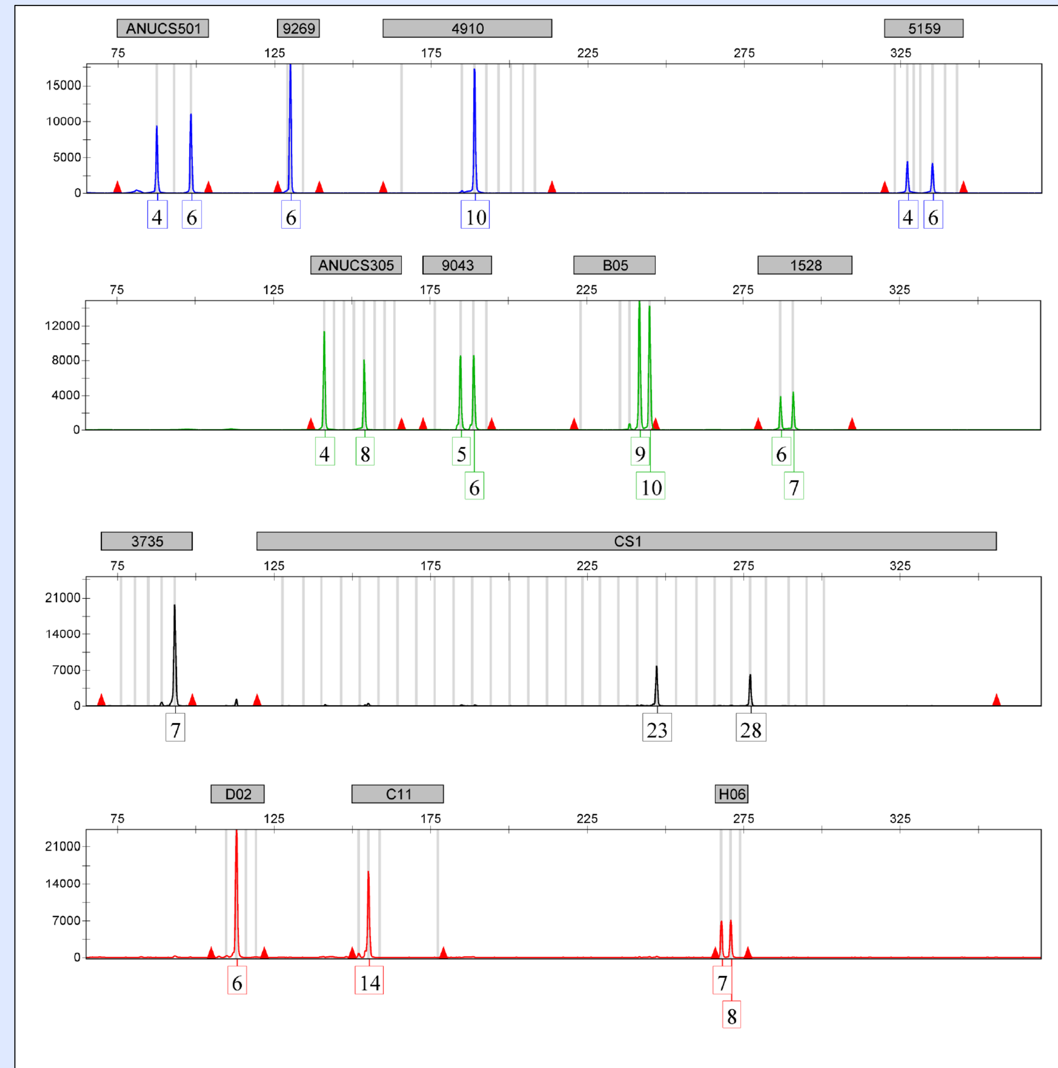

Fig. 1: Multiplex profile of 13 *Cannabis* STR loci using 0.5 ng of control template DNA (sample 1-D1).

Table 1: Characteristics of 13 *Cannabis* STR markers used in this study.

Marker	Dye	STR motif	Repeat type	Observed Alleles	[Primer] (µM)	Genbank accession no.
ANUCS501	6-FAM	(TTGTG)	Simple	4,5,6	0.10	KT203577-8
9269	6-FAM	(ATAA)	Simple	5.3,6,7	0.10	KX668131-2
4910	6-FAM	(AAGA)	Compound	4,10,14,15	0.20	KX668123-4
5159	6-FAM	(AGAT)	Simple	3,4,4.2,5.3,8,10	0.30	KX668125-7
ANUCS305	VIC	(TGG) (TGA) (GGG)	Simple	4,6,8,11	0.10	KT203571-3
9043	VIC	(TCTT)	Simple**	3,5,6	0.15	KX668128-30
B05	VIC	(TTG)	Simple	3,7,8,9,10	0.15	KT203581-2
1528	VIC	(ATTA)	Simple	6,7	0.30	KX668119-20
3735	NED	(TATG)	Simple	3,4,5,6,7	0.10	KX668121-2
CS1	NED	(ATCACC)*	Simple	10,12,13,16,17,23,24, 25,26,27,28,29,32	0.25	KT203586-90
D02	PET	(GTT)	Simple	6,7,8	0.15	KT203591-2
C11	PET	(TGG)x (TGA)y	Compound/indel	13,14,15,21	0.15	KT203583-5
H06	PET	(AAC) (GAC) (GAT) (AAT)	Simple	7,8,9	0.15	K203596-7

- *Cannabis* 13-loci multiplex STR system was optimized using the Type-IT Microsatellite PCR Kit (Qiagen)
- Optimal PCR conditions: Annealing Temp.= 57°C; 29 cycles
- Sequenced allelic ladder developed with 55 alleles across 13 STR loci
- All samples successfully amplified (N=101) under optimized conditions

Validation Studies

Table 2: Observed stutter ratios, peak height ratios, and inter-loci balance at each locus in the multiplex system for samples (N=25) amplified using 0.5ng of template DNA.

Marker	Stutter Ratio (Upper range)	Mean PHR	Inter-loci balance
ANUCS501	0.141	0.884	1.076
9269	0.018	0.694	1.019
4910	0.064	0.853	0.886
5159	0.038	0.823	0.500
ANUCS305	0.033	0.845	1.296
9043	0.027	0.889	1.063
B05	0.077	0.861	1.671
1528	0.021	0.838	0.617
3735	0.067	0.895	0.759
CS1	0.074	0.689	0.812
D02	0.048	0.861	1.561
C11	0.166	0.831	1.011
H06	0.149	0.820	0.728

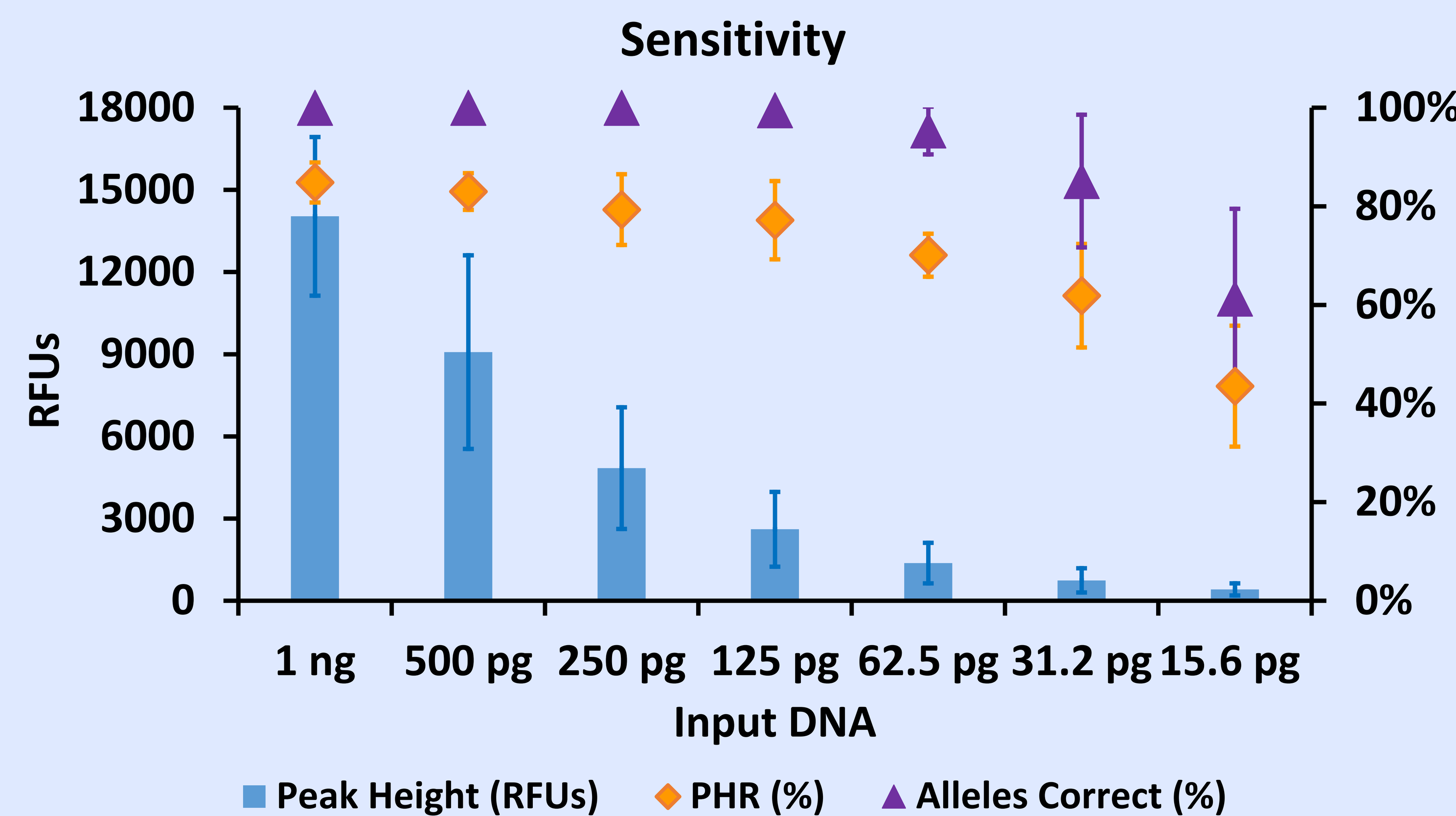

Fig. 2: Summary of sensitivity study from serially diluted single-source template DNA (n=5) ranging from 1ng to 20 pg. Error bars represent standard deviation.

- Average stutter percentage across all loci (2.14%) with a maximum upper range at marker C11 (16.6%)
- Mean Peak Height Ratio (PHR) was 83% across all loci
- Inter-loci balance range was 0.500 – 1.671

MATERIALS AND METHODS

Allele Sequencing For the tetranucleotide markers, at least two homozygous samples were selected for sequencing. PCR amplification and cycling sequencing was performed on the Veriti® Thermal Cycler (Applied Biosystems) using the BigDye® Direct Cycle Sequencing Kit (Applied Biosystems).

Validation Studies Validation studies included: sensitivity, species specificity, and genetic variation in a reference population. Additionally, internal validation studies were performed to provide detailed assessments of precision and accuracy, stutter ratios, peak height ratios (PHRs), inter-loci balance of the assay, and concordance with markers used in a previous multiplex.

Statistical Analysis For all STR markers, the number of multi-locus genotypes and the genotype sharing among samples were determined. For the reference population database (N=95), allele frequencies and parameters of forensic interest were estimated using the PowerStats v.12 software [8]. In addition, exact tests for Hardy-Weinberg equilibrium and linkage disequilibrium were performed on this reference population with the Genetic Data Analysis v.1.0 (GDA) software [9]. The *p* value for statistically significant differences was set at 0.05 levels.

CONCLUSIONS

- High quality profiles with template input as low as 0.13 ng
- None of the 13 STR markers cross-reacted with any of the studied species , except for *H. lupulus* (hops) which generated unspecific peaks
- STR success rates improved from previous multiplex (100% vs. 64%)
- Combined power of discrimination of the multiplex is 1 in 55 million
- No departures from Hardy-Weinberg or linkage equilibrium detected

REFERENCES

[1] C. Schield, C. Campelli, J. Sycalik, C. Randle, S. Hughes-Stamm, D. Gangitano, Identification and persistence of Pinus pollen DNA on cotton fabrics: A forensic application, Sci. Justice 56 (2016) 29-34. doi: 10.1016/j.scijus.2015.11.005.
[2] J.H. Bock, D.O. Norris, Forensic plant science, first ed., Elsevier Academic Press, London, 2016.
[3] D.N. Zaya, M.V. Ashley, Plant genetics for forensic applications, Methods Mol. Biol. 862 (2012) 35-52. doi: 10.1007/978-1-61779-609-8_4.
[4] L. Valverde, C. Lischka, S. Erlemann, E. de Meijer, M.M. de Pancorbo, H. Pfeiffer, S. Köhnemann, Nomenclature proposal and SNPSTR haplotypes for 7 new Cannabis sativa L. STR loci, Forensic Sci. Int. Genet. 13 (2014) 185-186. doi: 10.1016/j.fsigen.2014.08.002
[5] L. Valverde, C. Lischka, S. Scheiper, J. Nedele, R. Challis, M.M. de Pancorbo, H. Pfeiffer, S. Kohnemann, Characterization of 15 STR cannabis loci: nomenclature proposal and SNPSTR haplotypes, Forensic Sci. Int. Genet. 9 (2014) 61-65. doi: 10.1016/j.fsigen.2013.11.001.
[6] R. Houston, M. Birck, S. Hughes-Stamm, D. Gangitano, Evaluation of a 13-loci STR multiplex system for Cannabis sativa genetic identification, Int. J. Legal Med. 130 (2016) 635-647. doi: 10.1007/s00414-015-1296-x.
[7] R.A.L. Griffiths, M.D. Barber, P.E. Johnson, S.M. Gillbard, M.D. Haywood, C.D. Smith, J. Arnold, T. Burke, A.J. Urquhart, P. Gill, New reference allelic ladders to improve allelic designation in a multiplex STR system, Int. J. Legal Med. 111 (1998) 267-272.
[8] A. Tereba, Tools for analysis of population statistics., Profiles in DNA 3., Promega Corporation, 1999.
[9] P.O. Lewis, D. Zaykin, Genetic data analysis: Computer program for the analysis of allelic data, v. 1.0 (d16c), 2001. http://hydrodictyon.eeb.uconn.edu/people/plewis/downloads/gda-1.1.win32.zip (Accessed July 2016)

ACKNOWLEDGEMENTS

This work was supported by Award #2015-R2-CX-0030 (National Institute of Justice, Office of Justice Programs, U.S. Department of Justice). The opinions, findings, and conclusions or recommendations expressed in this presentation are those of the author(s) and do not necessarily reflect those of the Department of Justice.