Evaluation of a 13-loci STR Multiplex for *Cannabis sativa* Genetic Identification

Rachel Houston[1] B.S.,Matthew Birck[2] Ph.D, Sheree Hughes-Stamm[1] Ph.D, David Gangitano[1] Ph.D*
[1]Department of Forensic Science, College of Criminal Justice, Sam Houston State University, Huntsville, TX 77341
[2]New York Laboratory, U.S. Customs and Border Protection, U.S. Department of Homeland Security, Newark, NJ 07102

U.S. Customs and Border Protection

INTRODUCTION

Cannabis sativa L. is a plant cultivated worldwide as a source of fiber (hemp), medicine, and intoxicant [1,2]. In the United States, marijuana is the most commonly used illicit substance [3]. Consequently, marijuana is a highly trafficked drug to and within the United States by organized crime syndicates.

The development of a validated method using molecular markers, such as short tandem repeats (STRs) for the genetic identification of *C. sativa* will aid in the individualization of *Cannabis* samples as well as serve as an intelligence tool to link *Cannabis* cases (e.g., illegal traffic at the US-Mexico border).

As STRs are considered the gold standard for human identification, research has focused on the development of STR panels to identify marijuana plants [4]. In the United States, there have been attempts to create an STR database for *Cannabis* [5] as well as extensive research on a hypervariable STR marker, CS1 [6].

However, more comprehensive genetic tools need to be developed to provide a better insight into the genetic variation of marijuana. In addition, none of the previously published reports using *Cannabis* STR profiling have followed two important International Society of Forensic Genetics (ISFG) recommendations for the use of non-human DNA in forensic investigations : a) the use of sequenced allelic ladders for accurate designation of alleles and inter-laboratory STR profile sharing and b) relevant population and forensic parameters studied in a reference population database of *C. sativa* for random match probability estimations or verification of genetic relatedness [7].

This study expands upon the earlier work of Köhnemann *et al.,* which described a 15 STR multiplex for the individualization of marijuana [8]. We developed an accurate real-time PCR DNA quantification method for *C. sativa*, and evaluated a 13-loci STR multiplex method for genotyping marijuana following ISFG/SWGDAM guidelines (i.e., use of sequenced allelic ladder, sensitivity, species specificity).

MATERIALS AND METHODS

Sample Collection Marijuana samples (N=199) were obtained from 11 previously processed case sets at the U.S. Customs & Border Protection LSS Southwest Regional Science Center. A minimum of 10 specimens were randomly sampled from each case set. For collection, individual marijuana plant fragments (stem or flowers) were cut, with 10 mg of the plant tissue used for this study.

DNA Extraction DNA Extraction was then performed using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA) as per manufacturer's protocol. DNA samples were quantified by real-time PCR on a StepOne™ Real-Time PCR System (Applied Biosystems, Carlsbad, CA, USA) using SYBR Green PCR Master Mix (Applied Biosystems) and *C. sativa* specific primers (ANUCS304) .

STR Analysis STR profiling was conducted in a 13-loci multiplex format modified from a previous report [8]. Amplification of these markers was performed *via* PCR using the Type-IT Microsatellite PCR Kit (Qiagen) on the Eppendorf Master Cycler Gradient (Eppendorf, Hamburg, Germany). PCR products were run on the 3500 Genetic Analyzer (Applied Biosystems). A customized bin set was designed and an allelic ladder was included with each injection to ensure accurate genotyping.

RESULTS

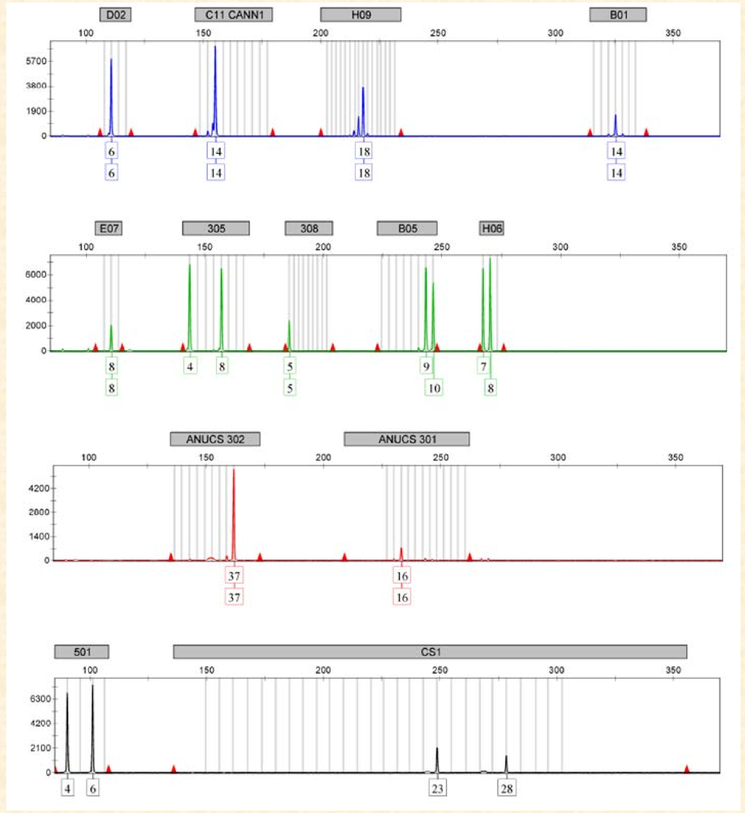

FIG. 1 – Multiplex profile of 13 Cannabis STR loci using 0.4ng of control template DNA .

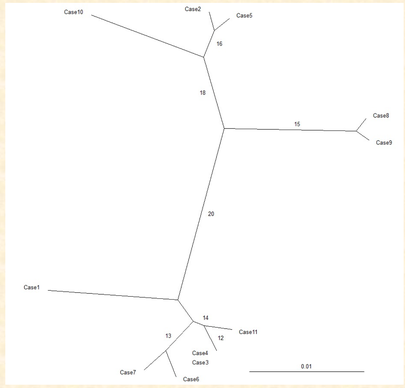

FIG. 2 – UPGMA tree depicting genetic distances among 11 Cannabis sample sets (N = 199) seized at the Mexico-US border. *Fs*t was set as genetic distance.

Marker	Dye	STR Motif	Type of Repeat	Observed alleles	Primer concentration (µM)	Genbank Accession No.
D02	FAM	(GTT)	Simple	6,7,8	0.04	KT203591-2
C11	FAM	$(TGA)_x(TGG)_y$	Compound/Indel	13,14,15,21	0.05	KT203583-5
H09	FAM	(GA)	Simple	11,13,16,18,19,21,23,24,25	0.08	KT203598-602
B01	FAM	$(GAA)_x(A)(GAA)_y$	Complex	11,13,14,15	0.09	KT203579-80
E07	VIC	(ACT)	Simple	7,8,9	0.30	KT203593-5
305	VIC	(TGG)	Simple	4,6,8,11	0.08	KT203571-3
308	VIC	(TA)	Simple	5,8,9,12	0.13	KT203574-6
B05	VIC	(TTG)	Simple	3,7,8,9,10	0.03	KT203581-2
H06	VIC	(ACG)	Simple	7,8,9	0.07	KT203596-7
501	NED	(TTGTG)	Simple	4,5,6	0.10	KT203577-8
CS1	NED	(CACCAT)	Simple	10,12,13,16,17,23,24,25,26,27,28,29,32	0.14	KT203586-90
302	PET	$(ACA)_x(ACA)_y(ACA)_z$	Compound	29,31,33,35,36,37	0.08	KT203569-70
301	PET	(TTA)	Simple	15,16,23,24,25	0.30	KT203566-8

TABLE 1 – Characteristics of 13 Cannabis STR markers used in this study.

MATERIALS AND METHODS

Allelic Ladder Design Fifty *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 [9,10].

Allele Sequencing Two to five homozygous samples representing the most common alleles were sequenced using the BigDye® Direct Cycle Sequencing Kit (Applied Biosystems). Sequences were then aligned and proofread using the Geneious Pro Software R8 (Biomatters, Auckland, New Zealand). Sequences were submitted to Genbank (Accession No. KT203566 to KT203602).

Validation Studies A sensitivity and specificity study were performed to access the dynamic range of the assay as well as cross-reactivity with non-Cannabis samples.

Statistical Analysis Phylogenetic analysis of different seizures was performed using the Unweighted Pair Group Method using Arithmetic averaging (UPGMA) method and coefficient of co-ancestry *Fst* as genetic distance were estimated with the Genetic Data Analysis (GDA) software. For the reference population database (N=97) allele frequencies and parameters of forensic interest were estimated using the PowerStats v. 12 software. In addition, population genetic statistics (number of alleles, observed heterozygosity, expected heterozygosity) as well as Hardy-Weinberg equilibrium and linkage disequilibrium tests were performed on this reference population. Null allele analysis was performed using the Genepop v4.2 Software; corrected allele frequencies were also reported.

CONCLUSIONS

- A real-time PCR method for C. sativa DNA quantitation was developed.
- Distinguishable DNA profiles were generated from 127 samples that yielded full STR profiles.
- Four duplicate genotypes within seizures were found.
- The combined power of discrimination of this multi-locus system is 1 in 70 million.
- The sensitivity of the multiplex STR system is 0.25 ng of template DNA.
- None of the 13 STR markers cross-reacted with any of the studied species, except for Humulus lupulus (hops).
- Phylogenetic analysis and case-to-case pairwise comparison of 11 cases using Fst as genetic distance revealed the genetic association of four groups of cases.
- Moreover, due to their genetic similarity, a subset of samples (N=97) was found to form a homogeneous population in Hardy-Weinberg and linkage equilibrium.

REFERENCES

1. Small E, Cronquist A (1976) A Practical and Natural Taxonomy for *Cannabis*. Taxon 25(4): 405-435.
2. Adams IB, Martin BR (1996) Cannabis: pharmacology and toxicology in animals and humans. Addiction 91:1585-1614
3. Administration SAaMHS (2014) Results from the 2013 National Survey on Drug Use and Health: Summary of National Findings. Substance Abuse and Mental Health Services Administration. http://www.samhsa.gov/data/sites/default/files/NSDUHresultsPDFWHTML2013/Web/NSDUHresults2013.pdf. Accessed April 29 2015
4. Howard C, Gilmore S, Robertson J, Peakall R (2008) Developmental validation of a Cannabis sativa STR multiplex system for forensic analysis. Journal of forensic sciences 53 (5):1061-1067. doi:10.1111/j.1556-4029.2008.00792.x
5. Mendoza MA, Mills DK, Lata H, Chandra S, ElSohly MA, Almirall JR (2009) Genetic individualization of Cannabis sativa by a short tandem repeat multiplex system. Analytical and bioanalytical chemistry 393 (2):719-726. doi:10.1007/s00216-008-2500-3
6. Shirley N, Allgeier L, LaNier T, Coyle HM Analysis of the NM01 Marker for a Population Database of Cannabis Seeds.
7. Linacre A, Gusmao L, Hecht W, Hellmann AP, Mayr WR, Parson W, Prinz M, Schneider PM, Morling N (2011) ISFG: recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. In: Forensic Sci Int Genet, vol 5. vol 5. 2010 Elsevier Ireland Ltd, Netherlands, pp 501-505. doi:10.1016/j.fsigen.2010.10.017
8. Kohnemann S, Nedele J, Schwotzer D, Morzfeld J, Pfeiffer H (2012) The validation of a 15 STR multiplex PCR for Cannabis species. International journal of legal medicine 126 (4):601-606. doi:10.1007/s00414-012-0706-6
9. Baechtel FS, Smerick JB, Presley KW, Budowle B (1993) Multigenerational amplification of a reference ladder for alleles at locus D1S80. Journal of forensic sciences 38 (5):1176-1182
10. Sajantila A, Puomilahti S, Johnsson V, Ehnholm C (1992) Amplification of reproducible allele markers for amplified fragment length polymorphism analysis. BioTechniques 12 (1):16, 18, 20-12