CHARACTERIZATION OF NEW CHLOROPLAST POLYMORPHISMS TO DETERMINE BIOGEOGRAPHICAL ORIGIN AND CROP TYPE OF *CANNABIS SATIVA* SAMPLES

Madeline G. Roman[1], BS, Michele Di Nunzio[2], MS, Rachel Houston[1], PhD, Bobby LaRue[1], PhD, and David Gangitano[1], PhD

[1]*Department of Forensic Science, Sam Houston State University, Huntsville, Texas, USA*

[2]*Forensic Genetics Laboratory – Legal Medicine Unit, Department of Medicine, University of Barcelona, Barcelona, Spain*

INTRODUCTION

Marijuana (*Cannabis sativa*) is the most commonly used illicit drug in the United States. Despite its schedule I classification by the federal government, 33 states have legalized its use for medicinal or recreational purposes. This state-specific legalization has created a new problem for law enforcement: preventing and tracking the diversion of legally-obtained cannabis to states where it remains illegal. In addition, illegal trafficking of the drug at the border with Mexico remains an issue for law enforcement agencies.

C. sativa crops can be broadly classified as marijuana (a drug containing the psychoactive chemical delta-9-tetrahydrocannabinol) or hemp (the non-drug form of the plant). Differentiation between crop types is important for forensic purposes. In addition, investigation of trafficking routes into and within the United States requires genetic association of samples from different cases/seizures, and knowing where the crop originated is extremely useful.

Previous studies have shown that the *ITS1*, *ITS2*, and *trnL-trnF* barcoding regions have sequences which are unique to *C. sativa* (1-3). While these regions are useful for identifying the species, they are not useful for determining the biogeographic origin or crop type. A study by Gilmore et al. (4) discovered five chloroplast and two mitochondrial DNA loci which varied among hemp and marijuana samples from different geographic origins. An additional study incorporated these seven markers in the creation of a DNA database for *C. sativa* samples (5). These studies suggest that organelle DNA can be used to distinguish marijuana and hemp (crop type) and crops grown in different parts of the world (biogeographic origin). However, it would be beneficial to increase the number of loci analyzed in order to better distinguish between *C. sativa* crops.

This project seeks to exploit sequence variations in *C. sativa* chloroplast DNA (cpDNA) to allow genetic determination of biogeographic origin, discrimination between drug-type and fiber-type samples (marijuana vs. hemp), and association between cases for *C. sativa* samples. By comparing published cpDNA sequences, 58 polymorphisms have been identified, as well as seven hotspot regions. Polymorphisms consist of homopolymeric repeats (hSTRs), insertion/deletions (INDELs), and single nucleotide polymorphisms (SNPs). In this project, hemp and marijuana samples from four countries (United States, Canada, Mexico, and Chile) were evaluated for informative polymorphisms in two cpDNA hotspot regions.

MATERIALS & METHODS

***In silico* Analysis of Published Genomes:** Published sequences of 6 *C. sativa* cultivars were obtained from the GenBank® database (NCBI) and aligned using Geneious Pro software R7.1.9 (Biomatters) and the Mauve genome alignment tool (6). Polymorphic loci were noted, and hotspot regions were identified where at least 3 polymorphisms existed within 1,600 bp.

Samples: DNA extracted from *C. sativa* samples using the DNeasy® Plant Mini Kit (QIAGEN) were obtained from collaborators. Sample groups consisted of marijuana from the USA-Mexico border (N=109), marijuana from Chile (N=21), hemp from Canada (N=11), hemp from the USA (N=7), and medical marijuana from Chile (N=4). DNA was quantified using a qPCR method previously developed and validated by Houston et al. (5).

Sanger Sequencing: To assess variability at the chosen polymorphic sites, sequencing was performed using the BigDye™ Direct Cycle Sequencing Kit (Thermo Fisher Scientific). Capillary electrophoresis was performed on a 3500 Genetic Analyzer (Thermo Fisher Scientific).

Fragment Analysis Development & Genotyping: Fragment analysis assays were designed to genotype hSTR and INDEL markers. Primers were designed using Primer3 (7) and AutoDimer softwares (8). Forward primers were labeled with 6-FAM™ fluorescent dye. PCR was carried out using Type-it Microsatellite PCR Master Mix (QIAGEN) with a target of 80 pg of DNA, and capillary electrophoresis was performed on a 3500 Genetic Analyzer (Thermo Fisher Scientific). Allelic ladders were designed to ensure accurate allele calling.

SNaPshot™ Assay Development & Genotyping: SNaPshot™ assays were designed to genotype SNP markers using the SNaPshot™ Multiplex Kit (Thermo Fisher Scientific). Primers were designed using Primer3 (7) and AutoDimer softwares (8). Single base extension primers were designed to anneal to the 20 bp sequence preceding the SNP site on the forward or reverse DNA strand. Capillary electrophoresis was performed on a 3500 Genetic Analyzer (Thermo Fisher Scientific).

Statistical Analysis: Genetic Data Analysis (GDA) software (9) was used to calculate genetic distance using the UPGMA method with coancestry identity. Principle component analysis was carried out using Past3 software (10).

RESULTS & DISCUSSION

Table 1. Hotspot regions in the chloroplast DNA of *C. sativa*. Seven hotspot regions were identified which may be useful for determining crop type and biogeographical origin.

Region	hSTRs	SNPs	INDELs	Total Polymorphisms	Size of Fragment
trnK-matK-trnK	2	1	1	4	1130 bp
rps16	1	3	0	4	800 bp
trnS-trnG	2	3	0	5	765 bp
ycf3	3	0	0	3	1580 bp
accD-psaI	1	2	0	3	308 bp
clpP	4	0	1	5	1118 bp
rpl32-trnL	3	1	1	5	417 bp

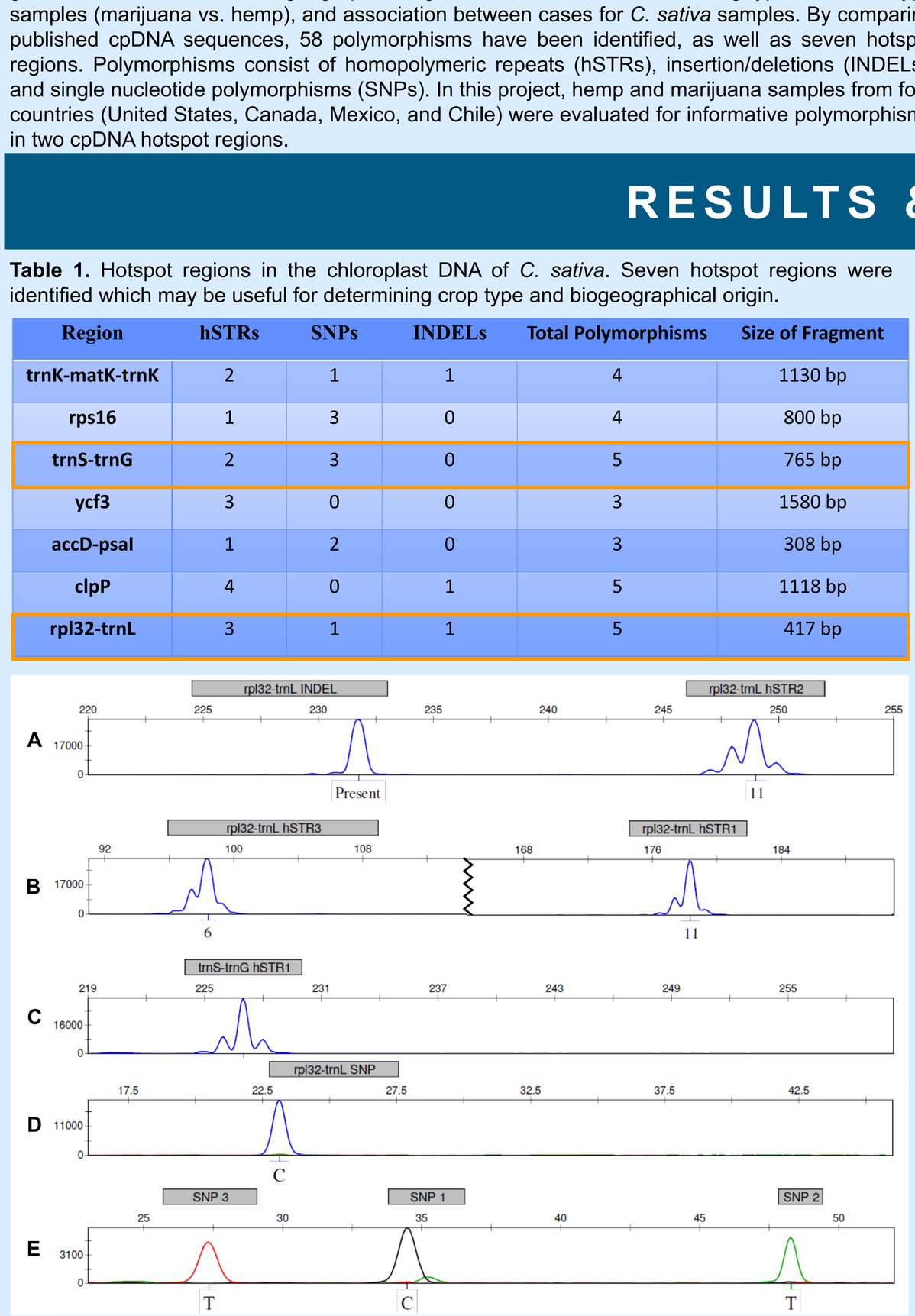

Figure 1. Example profiles of a marijuana sample. Fragment analysis: A) rpl32-trnL INDEL and hSTR2; B) rpl32-trnL hSTR3 and hSTR1; and C) trnS-trnG hSTR1. SNaPshot™ assays: A) rpl32-trnL SNP and B) trnS-trnG 3-plex SNP.

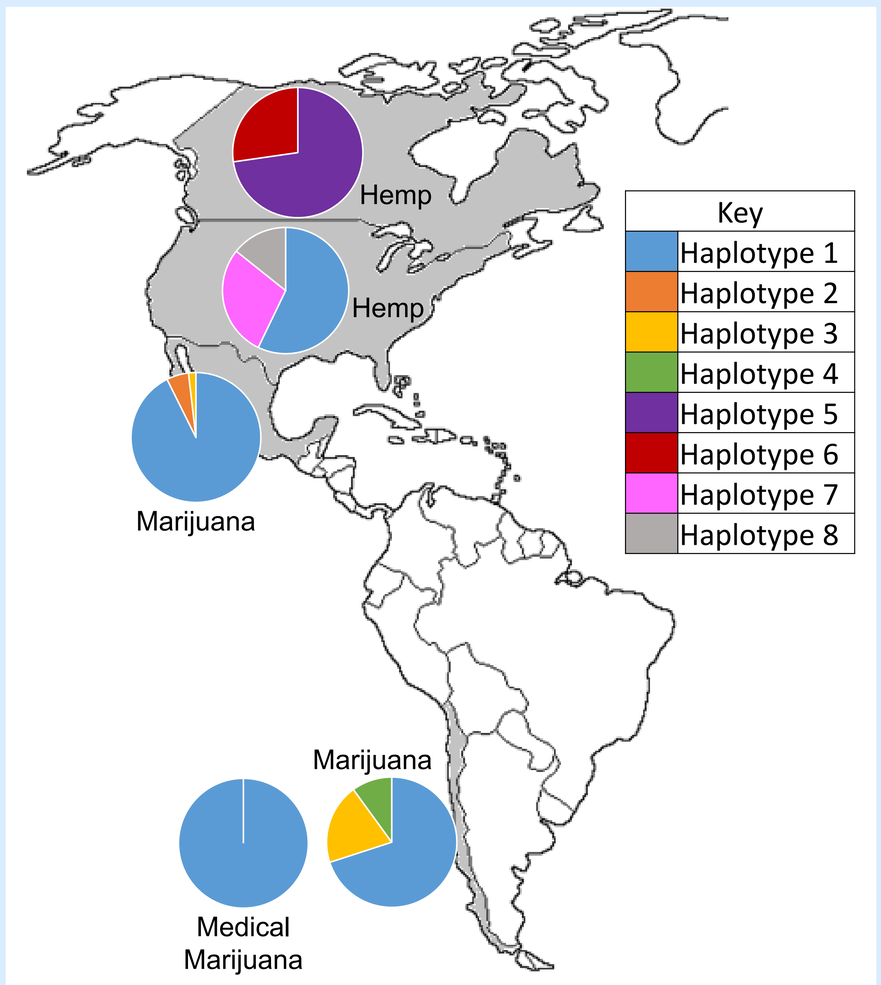

Figure 2. Map showing haplotype distribution among hemp and marijuana samples from four countries.

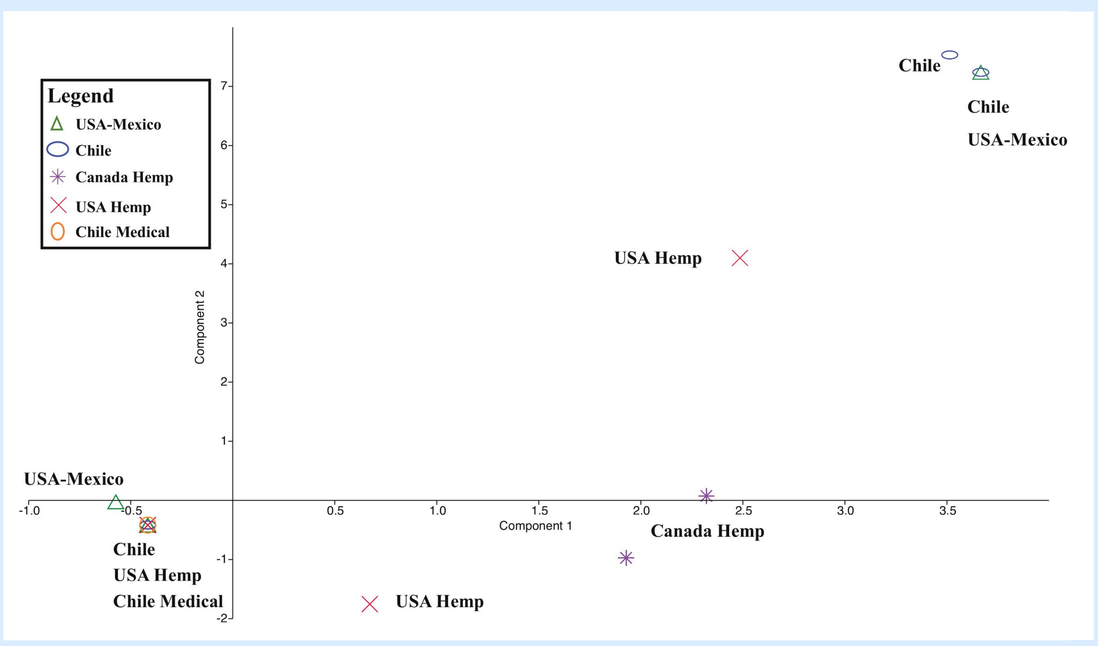

Figure 3. Principal component analysis plot showing some distinction between 5 sample groups by crop type and biogeographical origin.

CONCLUSIONS

- A total of 58 polymorphisms were discovered in the *C. sativa* chloroplast genome: 31 hSTRs, 23 SNPs, and 4 INDELs
  - 7 hotspot regions were identified
- Fragment analysis and SNaPshot™ assays were designed to genotype 2 hotspot regions (rpl32-trnL and trnS-trnG)
- 152 hemp and marijuana samples from 4 countries were characterized using the developed assays
- 8 haplotypes existed with haplotype 1 occurring in 81.6% of samples; haplotypes 2, 4, 5, 6, 7, and 8 were identified in only a single sample group
- Principle component analysis showed limited ability to distinguish between the 5 sample groups
- For the best results, a combination of these and other markers should be considered for distinguishing between samples of different crop types and biogeographical origins

ACKNOWLEDGEMENTS

Funding for this research was provided by the Department of Forensic Science at Sam Houston State University.

REFERENCES

1. Gigliano GS, Caputo P, Cozzolino S. Ribosomal DNA analysis as a tool for the identification of Cannabis sativa L. specimens of forensic interest. Science & Justice. 1997;37(3):171-4.
2. Gigliano SS. Preliminary data on the usefulness of internal transcribed spacer I (ITS1) sequence in Cannabis sativa L. identification. Journal of Forensic Science. 1999;44(3):475-7.
3. Linacre A, Thorpe J. Detection and identification of cannabis by DNA. Forensic science international. 1998;91(1):71-6.
4. Gilmore S, Peakall R, Robertson J. Organelle DNA haplotypes reflect crop-use characteristics and geographic origins of Cannabis sativa. Forensic science international. 2007;172(2):179-90.
5. Houston R, Birck M, LaRue B, Hughes-Stamm S, Gangitano D. Nuclear, chloroplast, and mitochondrial data of a US cannabis DNA database. International journal of legal medicine. 2018:1-13.
6. Aaron E. Darling, Bob Mau, and Nicole T. Perna. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss, and Rearrangement. PLoS One. 2010; 5(6):e11147.
7. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics. 2007;23(10):1289-91.
8. Vallone PM, Butler JM. AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques. 2004;37(2):226-31.
9. Lewis P. Genetic data analysis: computer program for the analysis of allelic data. Version 1.0 (d16c). http://lewis.eeb.uconn.edu/lewishome/software.html.2001.
10. Hammer Ø, Harper DA, Ryan PD. PAST: paleontological statistics software package for education and data analysis. Palaeontologia electronica. 2001;4(1):9.