Nuclear, Chloroplast, and Mitochondrial Data of a US Cannabis DNA Database

Rachel Houston BS*; Sheree Hughes-Stamm PhD; David Gangitano PhD

Department of Forensic Science, Sam Houston State University, Huntsville, TX 77340

MEMBER THE TEXAS STATE UNIVERSITY SYSTEM

INTRODUCTION

Cannabis sativa (marijuana) is a plant used for various purposes namely as an intoxicant, fiber, or medicine [1, 2]. Several genotyping methods have been suggested as a means of tracking and individualizing marijuana plants [3-6]. As with human identification, autosomal STR typing can be used as a means of individualizing cannabis samples.

In addition, biogeographical tracking could provide law enforcement insight on its trade and distribution patterns. Analysis of organelle DNA, including both mitochondrial and chloroplast DNA, has been shown to be a valuable tool in analyzing evolutionary and population diversity in plant species as it is inherited uniparentally [7-9].

In this work, a DNA database consisting of 496 samples was used to genotype both autosomal and organelle DNA. For this purpose, a previously validated 13-autosomal STR multiplex [5] was used to genotype 496 samples from four different sites: the US-Mexico border, Brazil, hemp seeds purchased in US, and Chile. For organelle typing, a previously reported multi-loci system from Gilmore et al. was modified and optimized to genotype five chloroplast and two mitochondrial markers [9]. For successful downstream organelle typing, a novel assay for the real-time PCR quantification of cannabis cpDNA using synthetic DNA standards was developed, optimized, and validated.

MATERIALS AND METHODS

DNA Collection Samples were analyzed from four different sites: 21 seizures at the US-Mexico border (N=408), Northeastern Brazil (N=8), hemp seeds purchased in US (N=30), and the Araucarian region of Chile (N=50).

Autosomal DNA Typing Autosomal DNA was quantified and genotyped in a 13-loci multiplex format according to Houston et al. [5,6].

Organelle DNA Quantitation Quantification of cpDNA was preformed via real-time PCR with a StepOne™ Real-Time PCR System (Thermo Fisher Scientific, South San Francisco, CA, USA) using SYBR Green PCR Master Mix (Thermo Fisher Scientific) and cannabis-specific chloroplast primers, Cscp001 (Integrated DNA Technologies, Coralville, IA, USA). Serial dilutions of Reconstituted synthetic DNA standards (1000 to 0.02 pg/µL) were used to generate a calibration curve. Validation studies included: sensitivity, specificity, precision and accuracy.

Organelle STR Typing Cannabis STR profiling was performed in a five – loci multiplex format consisting of previously published cannabis chloroplast (Cscp001, Cscp002, Cscp003, Cscp004) and mitochondrial (Csmt001) STR markers [9]. PCR amplification was performed using the Type-it Microsatellite PCR Kit (Qiagen) on a T100™ Thermal Cycler (Bio-Rad, Hercules, CA, USA).

RESULTS AND DISCUSSION

Organelle Typing

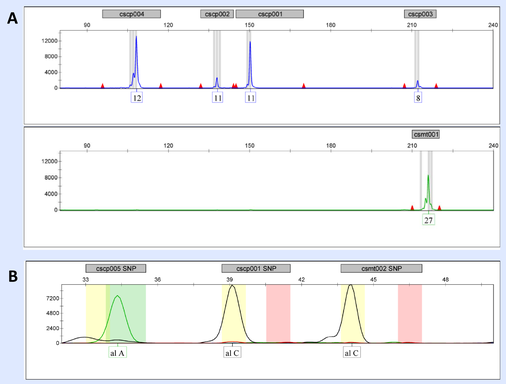

Fig. 1: Organelle haplotype **(A)** the homopolymer STR profile **(B)** the SNP profile.

- Cannabis organelle 5-loci multiplex STR system (**Fig. 1A**) was optimized using the Type-IT Microsatellite PCR Kit (Qiagen) using touchdown PCR method.
- Sequenced allelic ladder developed with 12 alleles across 5 STR loci.
- Cannabis organelle 3-loci SNP multiplex (**Fig. 1B**) was designed and optimized using the Type IT Microsatellite PCR Kit and SNaPshot™ Multiplex Kit (Thermo Fisher Scientific)
- Subsampling was performed and 134 samples were genotyped. Complete haplotypes (STRs and SNPs) were observed for 127 samples.
- Extensive haplotype sharing was observed; five distinguishable haplotypes were detected.
- Haplotype sharing was observed between the US border seizures, Brazil, and Chile while the hemp samples generated a distinct haplotype.

Statistical Analysis

Autosomal

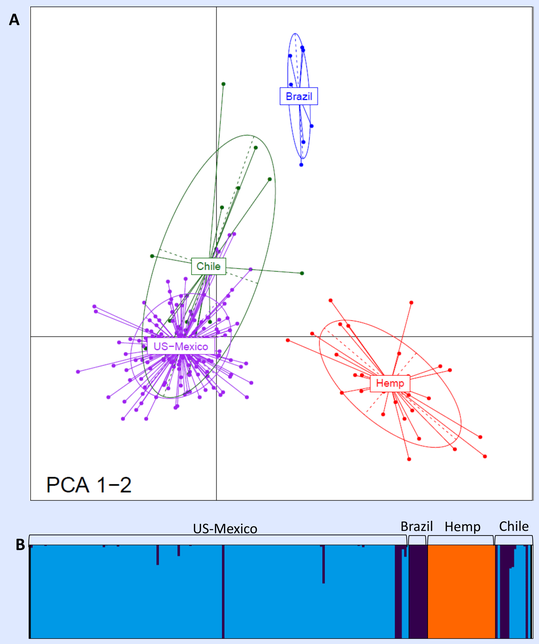

Fig. 2: Principal Component Analysis **(A)** Bayesian clustering with STRUCTURE **(B)** of autosomal genotypes from four cannabis datasets. The ellipses illustrate 95% inertia of each dataset while the dots represent individuals in the dataset. On the STRUCTURE bar plot, the colors depict the probability of assignment to each cluster (K=3).

Organelle

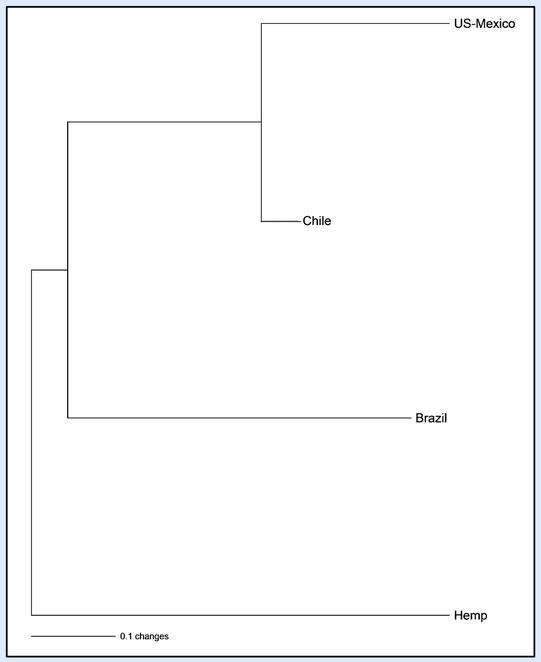

Fig. 3: Neighbor joining tree depicting genetic distances among four cannabis population sets using organelle haplotypes; coancestry as genetic distance.

MATERIALS AND METHODS

Organelle SNP Typing Cannabis SNP profiling was conducted in a three – loci multiplex consisting of previously published cannabis chloroplast (Cscp001 and Cscp005) and mitochondrial (Csmt002) SNP markers [9]. PCR amplification was performed using the Type-it Microsatellite PCR Kit (Qiagen) on a T100™ Thermal Cycler (Bio-Rad). The SBE (Single Base Extension) assay was performed on purified PCR products using the SNaPshot™ Multiplex Kit (Thermo Fisher Scientific) according to manufacturer's instructions [10].

Autosomal Statistical Analysis Case-to-case pairwise comparisons with F_{ST} as genetic distance were performed to determine a reference population from the 21 seizures ($p<0.05$). Phylogenetic analysis was assessed between the reference population, Brazil, Chile, and hemp samples using the neighbor joining method (Coancestry as genetic distance) with the Genetic Data Analysis (GDA) software [11]. Parsimony analysis was performed with PAUP* 4.0a [12]. Next, the STRUCTURE software was employed to evaluate the Bayesian clustering of genotypes from the four populations [13]. Finally, the R-based software, *Adegenet,* was used to visualize the data using Principal Component Analysis (PCA) [14].

Organelle Statistical Analysis Phylogenetic analysis was assessed between the four populations, and a distance matrix was calculated with the GDA software using the Neighbor Joining method (Coancestry as genetic distance) [11]. Next, PAUP* 4.0a was invoked to perform parsimony analysis [12].

CONCLUSIONS

- Statistical analysis
  - Reference population of 9 seizures (n=157)
  - Both autosomal and organelle markers could discern population sub-structure
  - Clear distinction between drug and hemp type samples
- Novel real-time qPCR method for quantifying cannabis cpDNA using synthetic standards developed and validated
- Organelle typing: two analytical methods were modified and optimized
  - Homopolymer STR pentaplex
  - SNP triplex with SNaPshot™ (Thermo Fisher Scientific)

REFERENCES

[1] E. Small, A. Cronquist, A Practical and Natural Taxonomy for Cannabis, Taxon 25(4) (1976) 406-435.
[2] I.B. Adams, B.R. Martin, Cannabis: pharmacology and toxicology in animals and humans, Addiction (Abingdon, England) 91(11) (1996) 1585-614.
[3] C. Howard, S. Gilmore, J. Robertson, R. Peakall, Developmental validation of a Cannabis sativa STR multiplex system for forensic analysis, J of Forensic Sciences 53(5) (2008) 1061-7 doi:10.1111/j.1556-4029.2008.00792.x.
[4] S. Kohnemann, J. Nedele, D. Schwotzer, J. Morzfeld, H. Pfeiffer, The validation of a 15 STR multiplex PCR for Cannabis species, Int J of Leg Med 126(4) (2012) 601-6 doi:10.1007/s00414-012-0706-6.
[5] R. Houston, M. Birck, S. Hughes-Stamm, D. Gangitano, Evaluation of a 13-loci STR multiplex system for Cannabis sativa genetic identification, Int J of Leg Med 130(3) (2015) 635-647 doi:10.1007/s00414-015-1296-x.
[6] R. Houston, M. Birck, S. Hughes-Stamm, D. Gangitano, Developmental and internal validation of a novel 13 loci STR multiplex method for Cannabis sativa DNA profiling, Legal Med (Tokyo, Japan) 26 (2017) 33-40 doi:10.1016/j.legalmed.2017.03.001.
[7] S. Dumolin-Lapegue, M.H. Pemonge, R.J. Petit, An enlarged set of consensus primers for the study of organelle DNA in plants, Mol Ecology 6(4) (1997) 393-7.
[8] M. Kohjyouma, I.J. Lee, O. Iida, K. Kurihara, K. Yamada, Y. Makino, S. Sekita, M. Satake, Intraspecific Variation in Cannabis sativa L. Based on Intergenic Spacer Region of Chloroplast DNA, Biol & Pharmaceutical Bull 23(6) (2000) 727-730 doi:10.1248/bpb.23.727.
[9] S. Gilmore, R. Peakall, J. Robertson, Organelle DNA haplotypes reflect crop-use characteristics and geographic origins of Cannabis sativa, Forensic Sci Int 172(2-3) (2007) 179-90 doi:10.1016/j.forsciint.2006.10.025.
[10] ABI PRISM® SNaPshot™ Multiplex Kit Protocol 43233578, in: A. Biosystems (Ed.) 2010.
[11] P. Lewis, D. Zaykin, Genetic data analysis: Computer program for the analysis of allelic data, 2001.
[12] D. Swofford, PAUP*: phylogenetic analysis using parsimony (* and other methods). Version 4. Sunderland, MA, Sinauer Associates, 2003.
[13] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155(2) (2000) 945-59.
[14] T. Jombart, adegenet: a R package for the multivariate analysis of genetic markers, Bioinformatics 24(11) (2008) 1403-5 doi:10.1093/bioinformatics/btn129.

ACKNOWLEDGEMENTS

This work was supported by Award #2015-R2-CX-0030, awarded by the 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.