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, USA

FSF Emerging Forensic Scientist Award
Paper Presentation

NIJ Disclosure

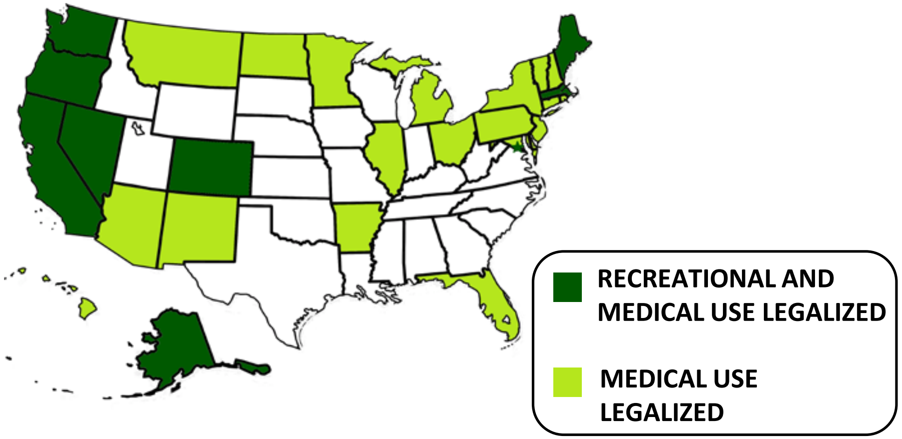
This study was partially funded by a Graduate Research Fellowship Award #2015-R2-CX-0030 (National Institute of Justice, Office of Justice Programs, U.S. Department of Justice). The opinions, findings, conclusions, or recommendations expressed in this presentation are those of the authors and do not necessarily reflect those of the National Institute of Justice.

Disclosure

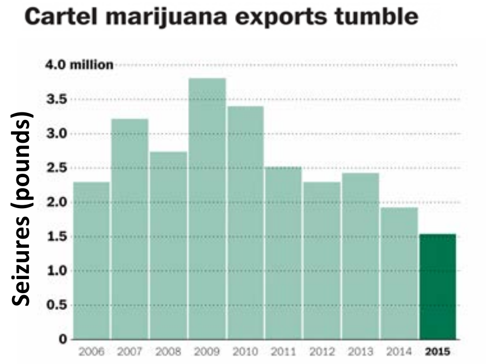
There is no real or apparent conflicts of interest related to the content of this presentation.

Statement of the Problem

Marijuana legalization

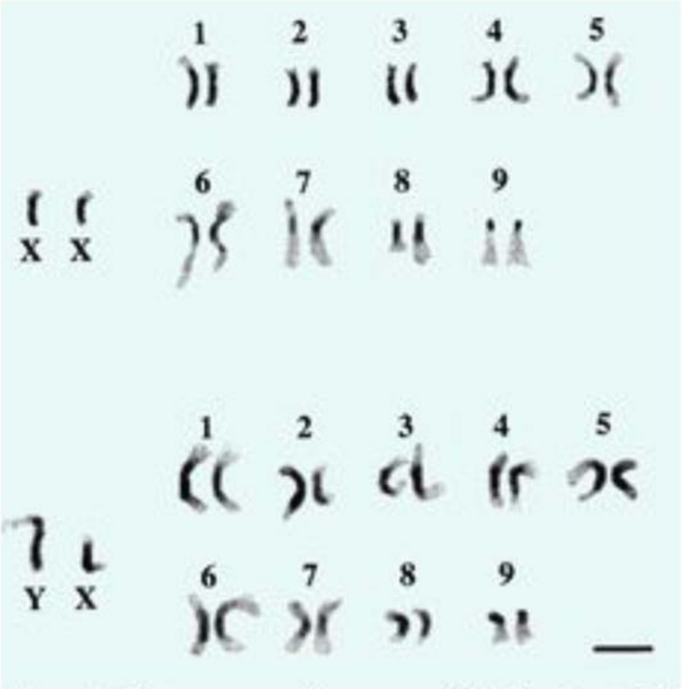


Illegal trafficking



Marijuana Background

- Family: Cannabaceae
- Dioecious plant
- Diploid genome: 2n = 20
    - 9 pairs of autosomes
    - Pair of sex chromosomes
- Draft genome complete[1]
    - 818 Mb for female (haploid)
    - 843 Mb for male (haploid)

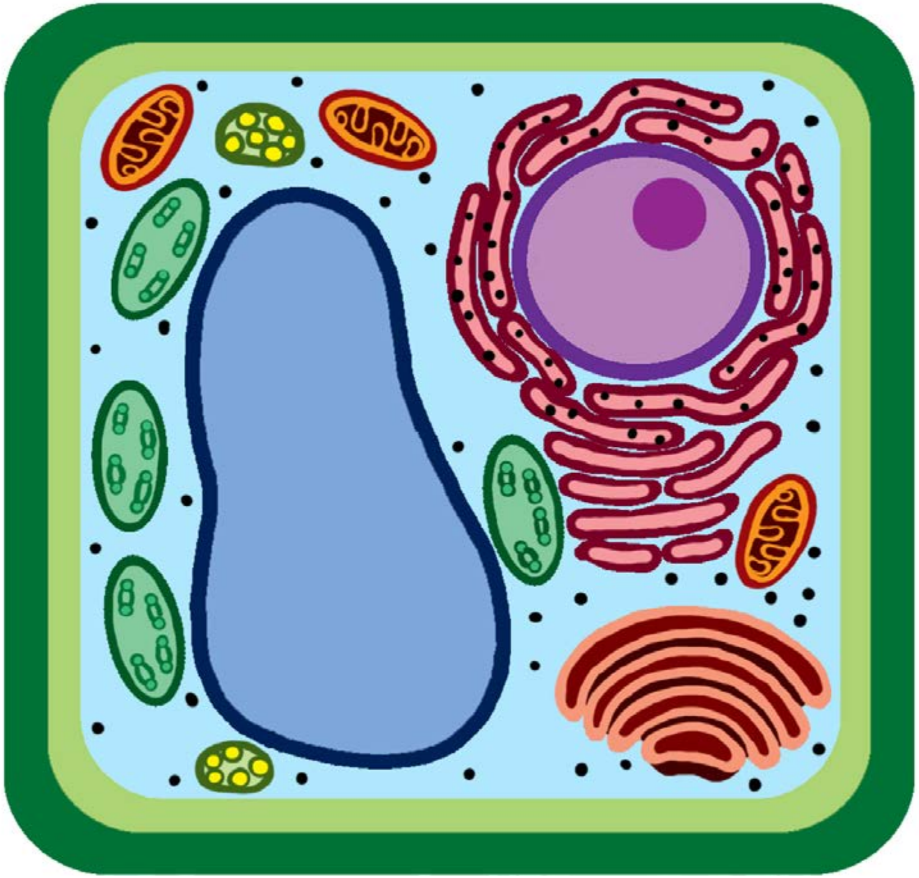


Cannabis Chromosome Karyogram by Sakamoto, et al.
The Japan Mendel Society, 1998

[1] van Bakel, Stout et al. 2011)

Organelle DNA

- Uniparentally inherited
- Circular genome
- Chloroplast
  - 153,871 bp[2]
- Mitochondria
  - 415,499 bp[3]

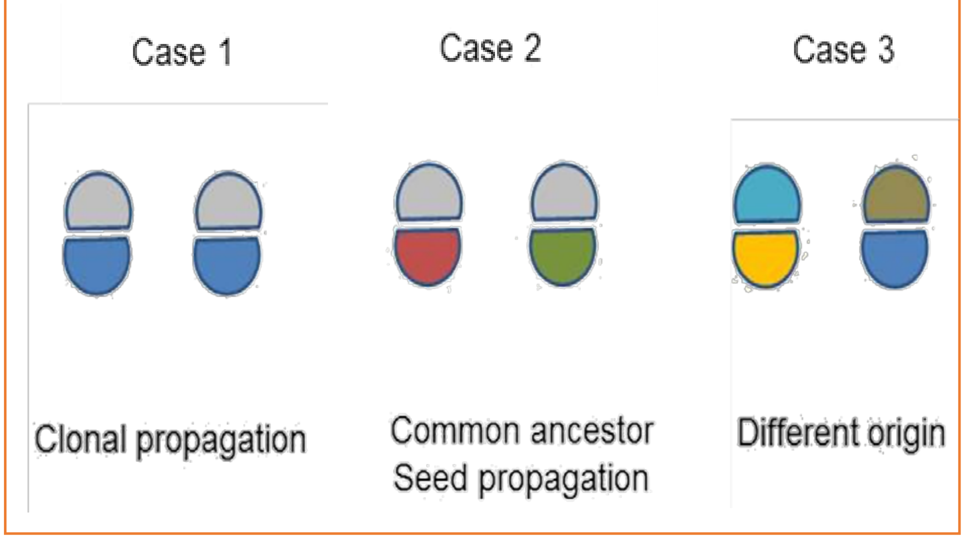


[2] Vergara, White et al. 2016
[3] Vergara, White et al. 2016

Goal of Research

- Build a database of DNA profiles
- Autosomal genotyping
    - Short tandem repeats
    - Individualizing plants
    - Genetic relatedness
- Organelle genotyping
    - Haplotypes
    - Biogeographical origin

DNA typing

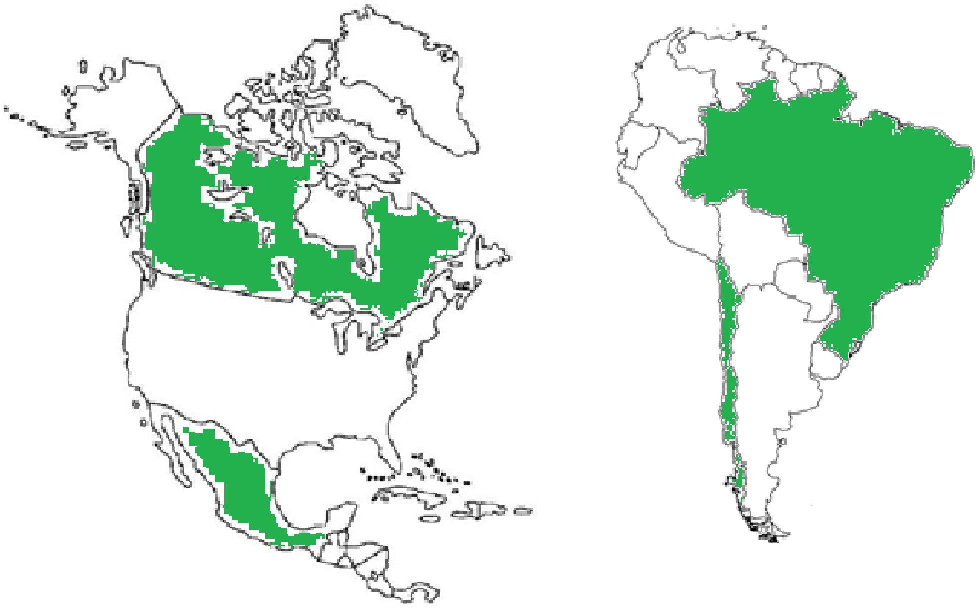


Materials and Methods

- Samples (4 sources)
- DNA Extraction (DNeasy® Plant Mini Kit)
- Autosomal DNA Typing
- Organelle DNA Typing
- Statistical Analysis

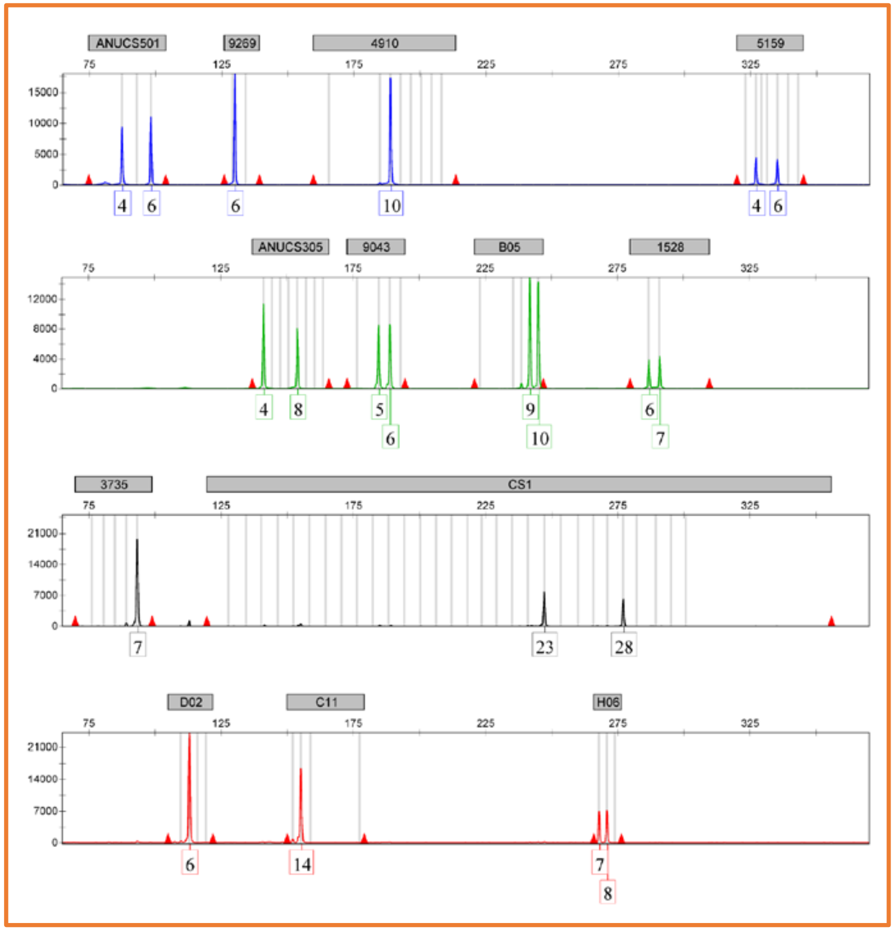
Samples

- US-Mexico Border:
  - 21 seizures
  - *N* = 422 samples
- Chile:
  - 10 seizures
  - *N* = 50 samples
- Brazil:
  - *N* = 8 samples
- Hemp Seeds:
  - Purchased in US (grown in Canada)
  - *N* = 30 samples



Autosomal DNA Typing

- Nuclear quantitation[4]
- Cannabis STR profiling using previously validated STR method[5]



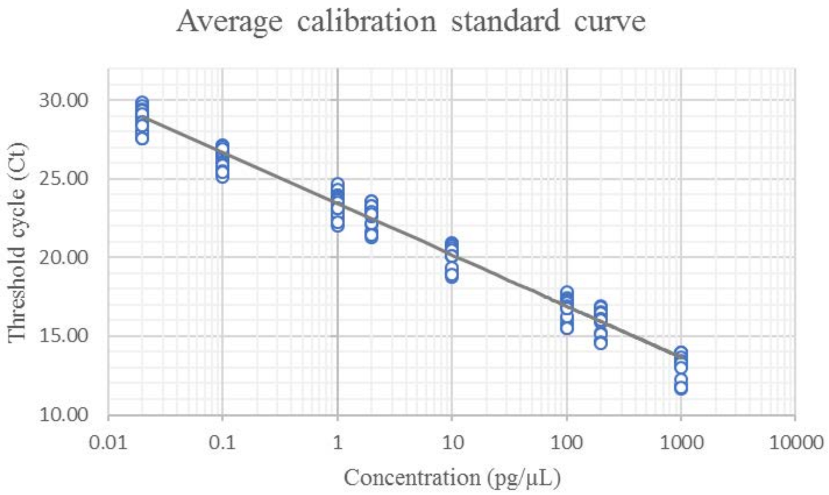
---

[4] Houston et al. 2016

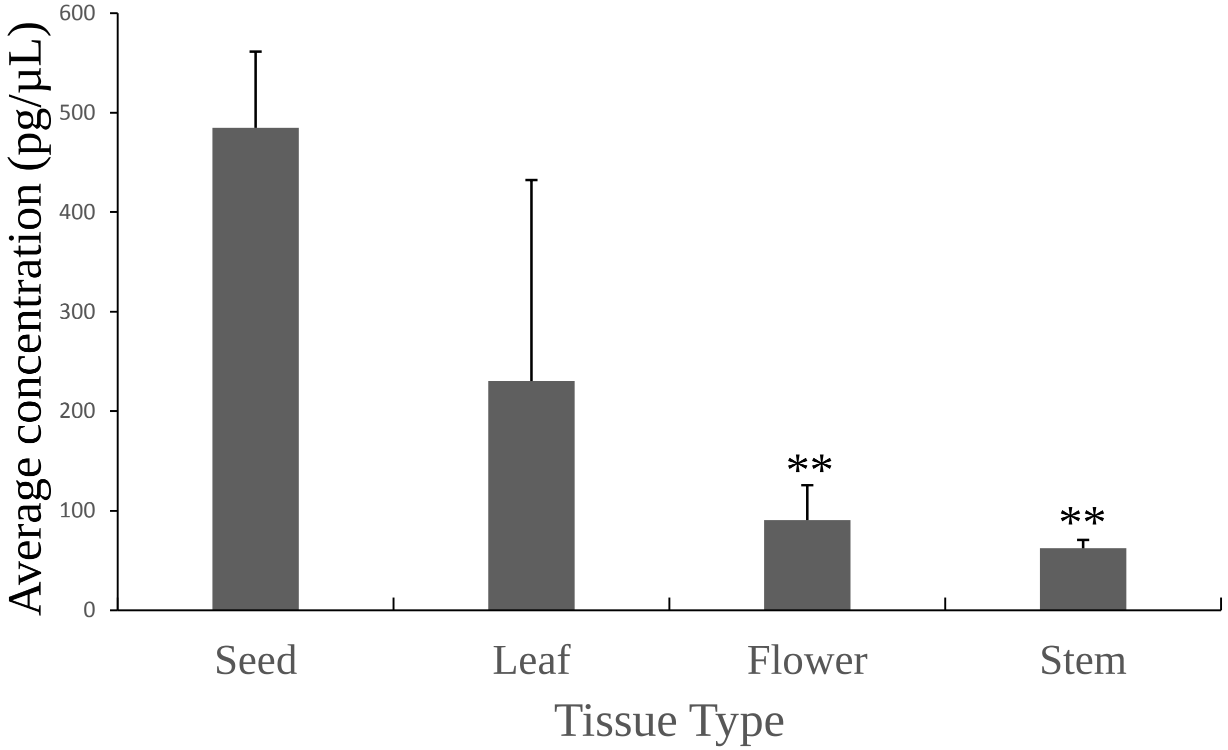
[5] Houston et al. 2017

Organelle DNA Quantitation

- Synthetic DNA standards:
  - Two complementary, PAGE-purified synthetic oligonucleotides
  - Cscp001 region of *C. sativa*
  - Diluted to generate standards 1000 – 0.02 pg/µL
- Quantitation via real-time PCR
- Validation studies:
  - Sensitivity
  - Specificity
  - Precision and accuracy



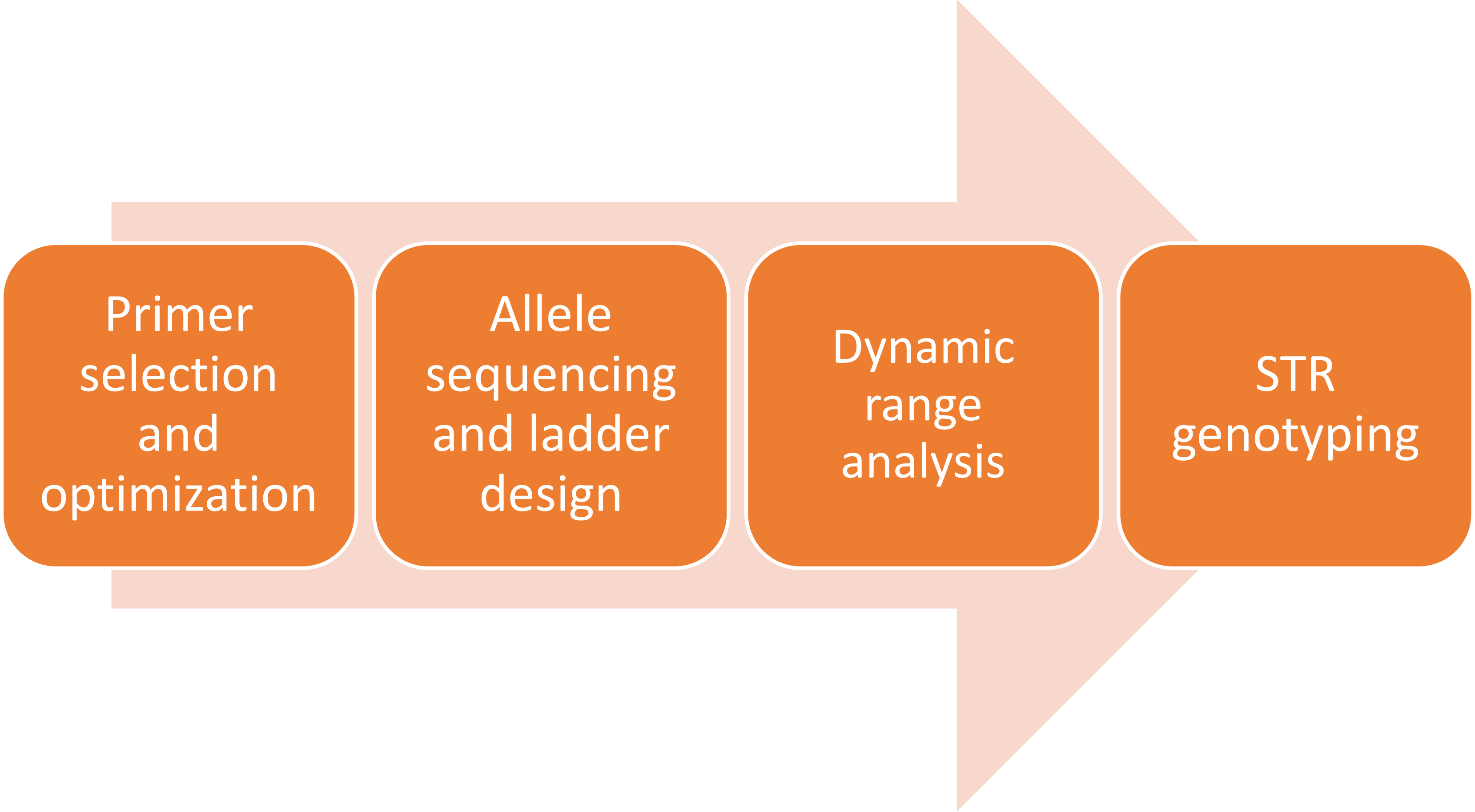
Relative cpDNA Quantitation



Error bars represent standard deviations.

** p-value < 0.01 when compared to seed tissue.

Organelle STR Typing

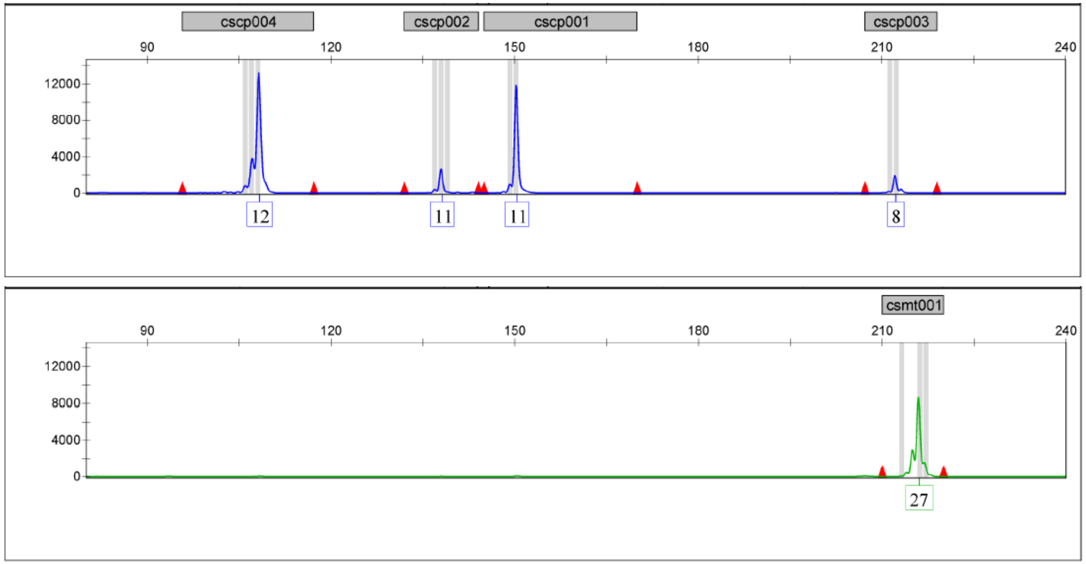


- Primer selection and optimization
- Allele sequencing and ladder design
- Dynamic range analysis
- STR genotyping

**** Modified from a previous study Gilmore et al. 2007**

Organelle STR Typing

Locus	Primer Reference	Region of DNA
STR based		
Cscp001	[6]Gilmore et al. 2007	*trn*L – *trn*F
Cscp002	[6]Gilmore et al. 2007	*rbc*L – orf106
Cscp003	[7]Weising et al. 1999	ccmp2 (5’ to *trn*S)
Cscp004	[7]Weising et al. 1999	ccmp6 (orf77 – orf82)
Csmt001	[6]Gilmore et al. 2007	*nad* 4 exon 3 to exon 4



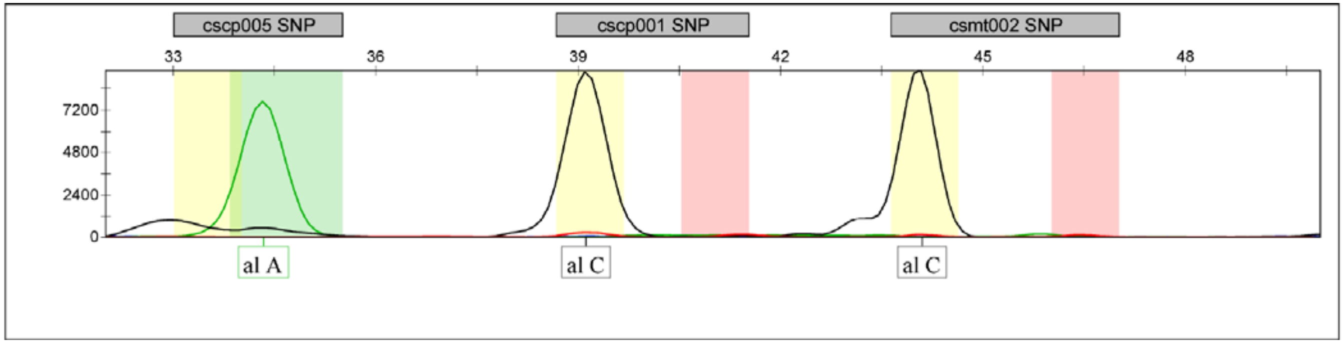
Organelle SNP Typing

- Primer design and optimization
- Allele confirmation via sequencing
- SNP genotyping (SNaPshot)

** Modified from a previous study Gilmore et al. 2007

Organelle SNP Typing

Locus	Primers	Primer Reference	Region of DNA
SNP based			
Cscp001	F: 5’ – tccctctatccccaaaaagg – 3’ R: 5’ – attgcacattgggattcctt – 3’ SBE F: 5’ – ttttttttttttacaaccggacctgaatgacc – 3’	This study	*trnL* – *trn*F
Cscp005	F: 5’ – tccactgccttgatccactt – 3’ R: 5’ – ccctctagacttagctgctct – 3’ SBE R: 5’ – cttttatcttgtctaaaattgaaat – 3’	This study	*trn*H – *trn*K
Csmt002	F: 5’ – tgtgcgaagagtgcgttatg – 3’ R: 5’ – acttcactcgctaggggatg – 3’ SBE F: 5’ – tttttttttttttttttttatgacctgtggccgcctg – 3’	This study	*nad* 5 exon 4 to exon 5



Statistical Analysis – Autosomal

- Number of multi-locus genotypes and genotype sharing determined
- Determine reference population
  - Phylogenetic analysis (GDA) and parsimony analysis (*PAUP*4.0a*)
  - Case-to-case comparisons with F_{ST}
- Phylogenetic analysis between reference population, Brazil, Chile, and hemp
  - Neighbor Joining method (GDA) and parsimony analysis (*PAUP*4.0a*)
  - Bayesian clustering with *STRUCTURE* software
  - Principal component analysis with *R* package, *Adegenet*

Statistical Analysis – Organelle

- Due to less variability, subsampling of 134 samples
- Number of haplotypes and haplotype sharing
- Phylogenetic analysis between reference population, Brazil, Chile, and hemp
  - Neighbor Joining method (GDA)
  - Parsimony analysis (*PAUP*4.0a*)

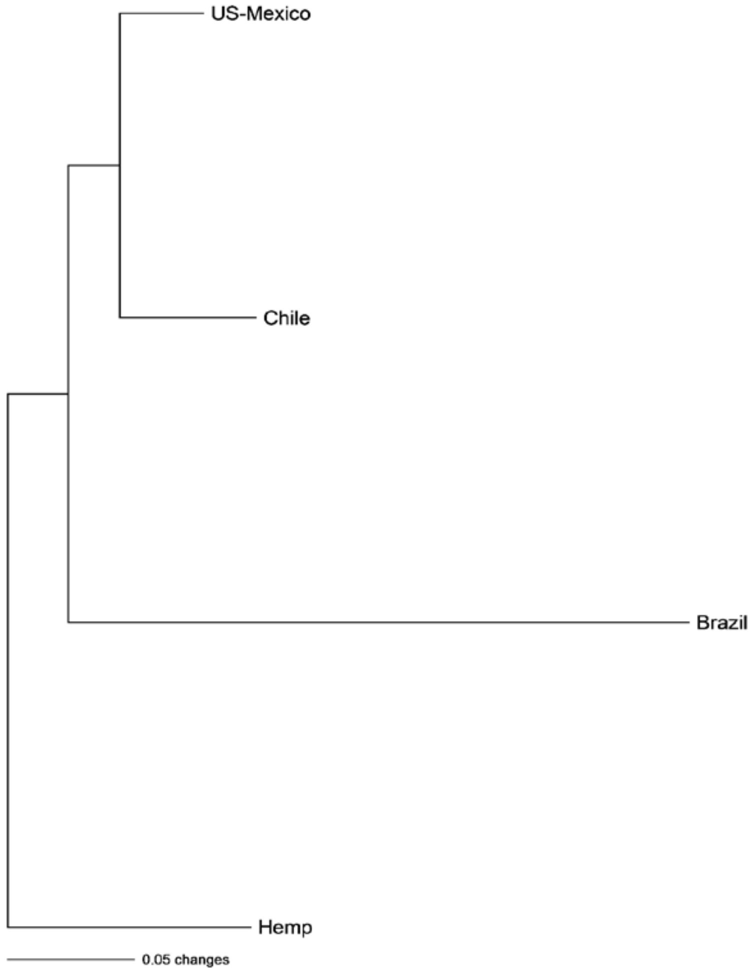
Autosomal – Genotypes

Source	Sample Number	Partial Profiles	Mixed Profiles	Full Profiles	Unique Genotypes	Duplicate Genotypes
US-Mexico	422	23	32	367	326	18 (41 samples)
Brazil	8	2	0	6	6	0
Chile	50	18	0	32	4	7 (28 samples)
Hemp	30	9	1	20	20	0
	510	52	33	425	356	25 (69 samples)

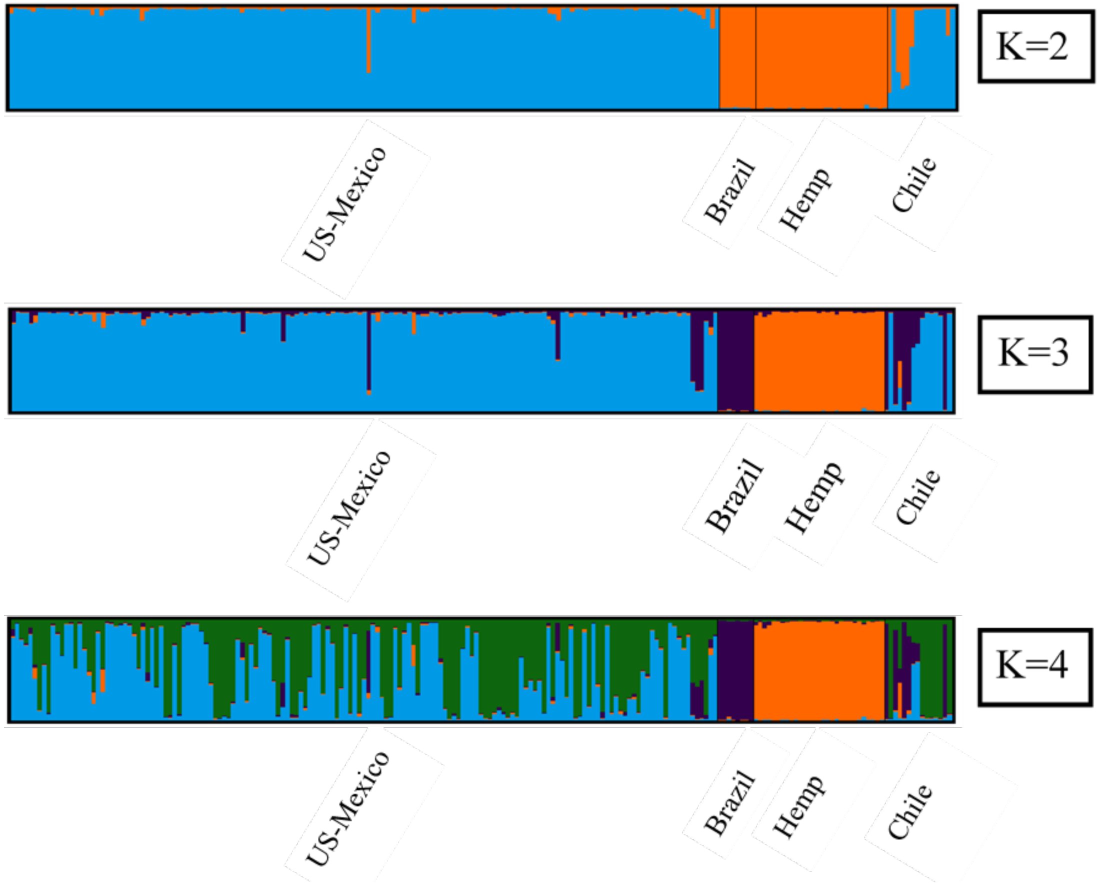
- Duplicates from same seizure or same plant
- Hypothesized that 9 out of 10 Chilean seizures clonally propagated

Autosomal – Phylogenetic Analysis

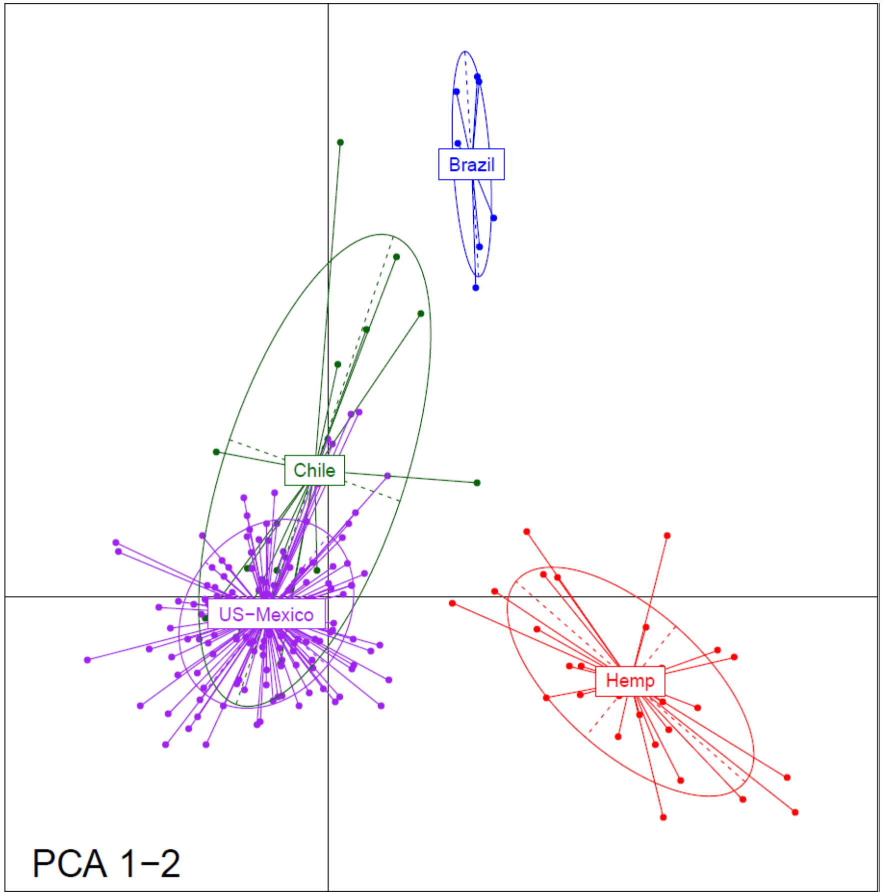
- 9 seizures (*N*=157 samples) formed a reference population for US-Mexico
  - $F_{ST} = 0$



Autosomal – Bayesian Clustering



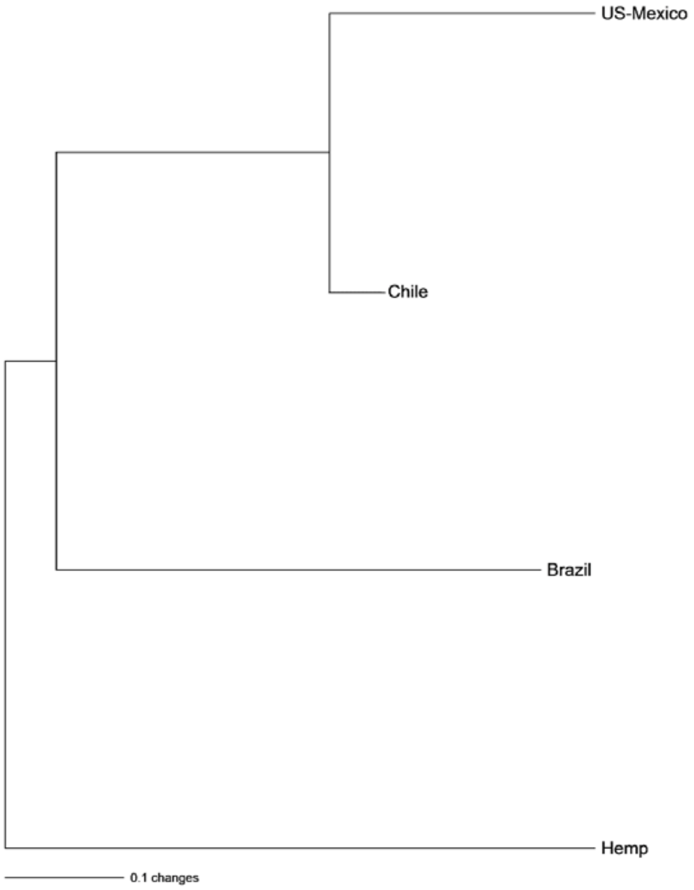
Autosomal – PCA



- Genetic relatedness between drug types
- Distinction from fiber type

Organelle – Statistical Analysis

- 127 out of 134 samples yielded full haplotypes
- Extensive haplotype sharing
- Five distinguishable haplotypes
- Haplotype sharing between US border seizures, Brazil, and Chile
- Distinct haplotype for hemp



Conclusions

1. 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
2. Novel real-time PCR method for quantifying cannabis cpDNA using synthetic standards developed and validated
3. Organelle typing: two analytical methods were modified and optimized
   - Homopolymer STR pentaplex
   - SNP triplex with SNaPshot™ (Thermo Fisher Scientific)

References

1. van Bakel H, Stout J, Cote A et al (2011) The draft genome and transcriptome of cannabis sativa http://genomebiology.Com/2011/12/10/r102. Genome Biology 1465-6906: R102
2. Vergara D, White KH, Keepers KG, Kane NC (2016) The complete chloroplast genomes of *Cannabis sativa* and *Humulus lupulus*. Mitochondrial DNA Part A, DNA mapping, sequencing, and analysis 27: 3793-3794. https://doi.org/10.3109/19401736.2015.1079905
3. White KH, Vergara D, Keepers KG, Kane NC (2016) The complete mitochondrial genome for *Cannabis sativa*. Mitochondrial DNA Part B 1: 715-716. https://doi.org/10.1080/23802359.2016.1155083
4. Houston R, Birck M, Hughes-Stamm S, Gangitano D (2015) Evaluation of a 13-loci STR multiplex system for *Cannabis sativa* genetic identification. Int J of Legal Med 130: 635-647. https://doi.org/10.1007/s00414-015-1296-x
5. Houston R, Birck M, Hughes-Stamm S, Gangitano D (2017) Developmental and internal validation of a novel 13 loci STR multiplex method for *Cannabis sativa* DNA profiling. Legal Med (Tokyo, Japan) 26: 33-40. https://doi.org/10.1016/j.legalmed.2017.03.001
6. Gilmore S, Peakall R, Robertson J (2007) Organelle DNA haplotypes reflect crop-use characteristics and geographic origins of *Cannabis sativa*. Forensic Sci Int 172: 179-190. https://doi.org/10.1016/j.forsciint.2006.10.025
7. Weising K, Gardner RC (1999) A set of conserved pcr primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42: 9-19

Acknowledgements

- American Academy of Forensic Sciences
- NIJ Award Number #2015-R2-CX-0030
- Sam Houston State University
- Dr. David Gangitano
- Dr. Sheree Hughes-Stamm

FSF Emerging Forensic Scientist Award
Paper Presentation

Questions

Rachel Houston

rmh034@shsu.edu