7th QIAGEN Investigator Forum, May 2018;
San Antonio, TX, USA

Innovative molecular approaches for *Cannabis sativa* DNA typing

Rachel Houston, PhD; Carrie Mayes, BS; Jonathan L. King, MS; Sheree R. Hughes-Stamm, PhD; David Gangitano, PhD

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

UNT HEALTH SCIENCE CENTER

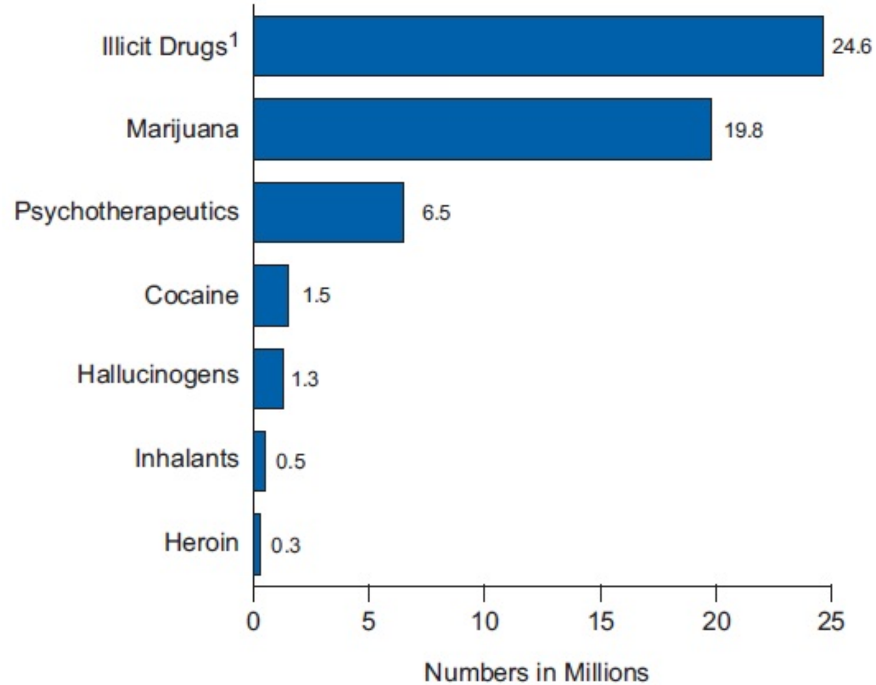
Sam Houston State University

Role of funding

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.

Significance

- Marijuana is the most commonly used illicit drug in United States

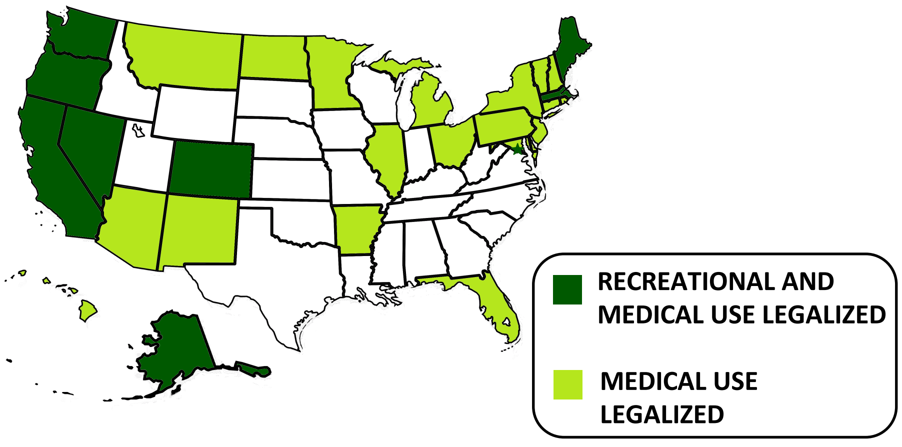


[1] Illicit Drugs include marijuana/hashish, cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically.

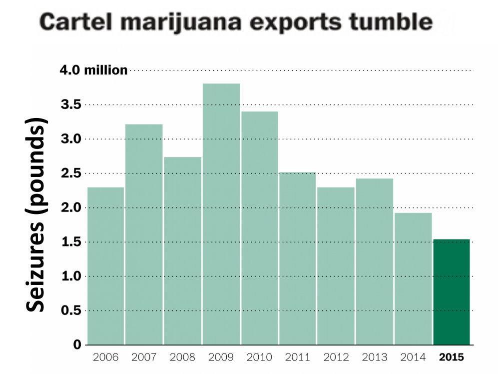
Statistics of Drug Use in United States (U.S. Department of Health and Human Services, 2013)

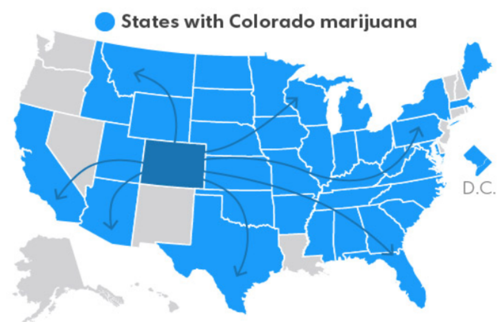
Statement of the problem

Marijuana legalization



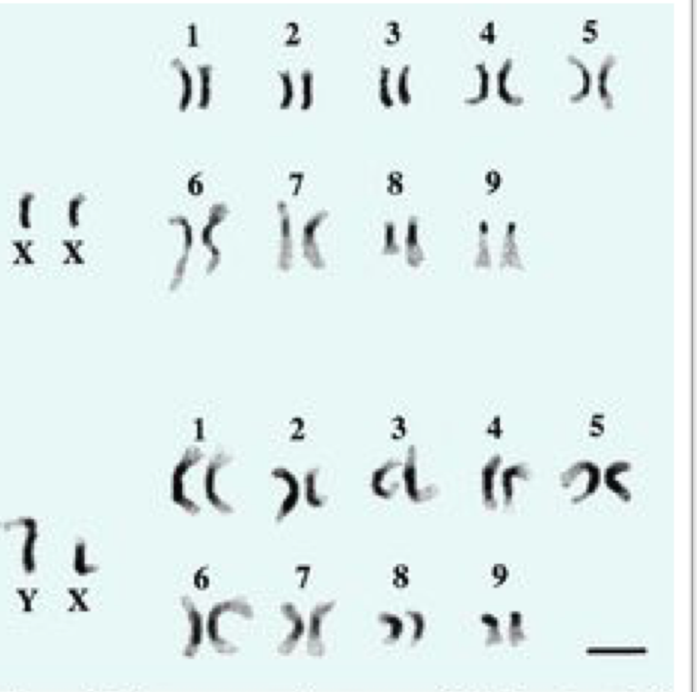
Illegal trafficking





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)

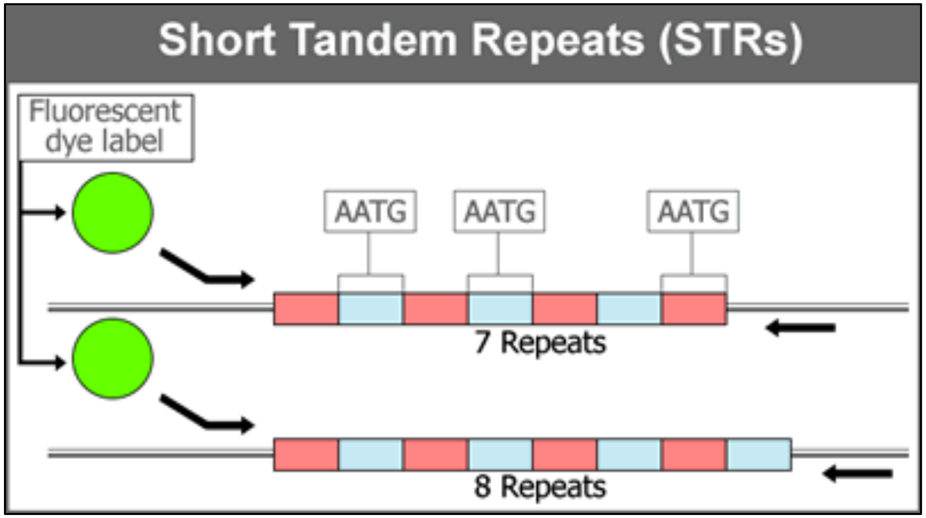


Cannabis Chromosome Karyogram by Sakamoto, et al.
The Japan Mendel Society, 1998

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

DNA based individualization

- **Autosomal short tandem repeat (STR)**
  - Gold standard in human identification
- Chloroplast DNA (cpDNA)
- Mitochondrial DNA (mtDNA)



http://rosalind.info/media/microsatpcr2.gif

Limitations

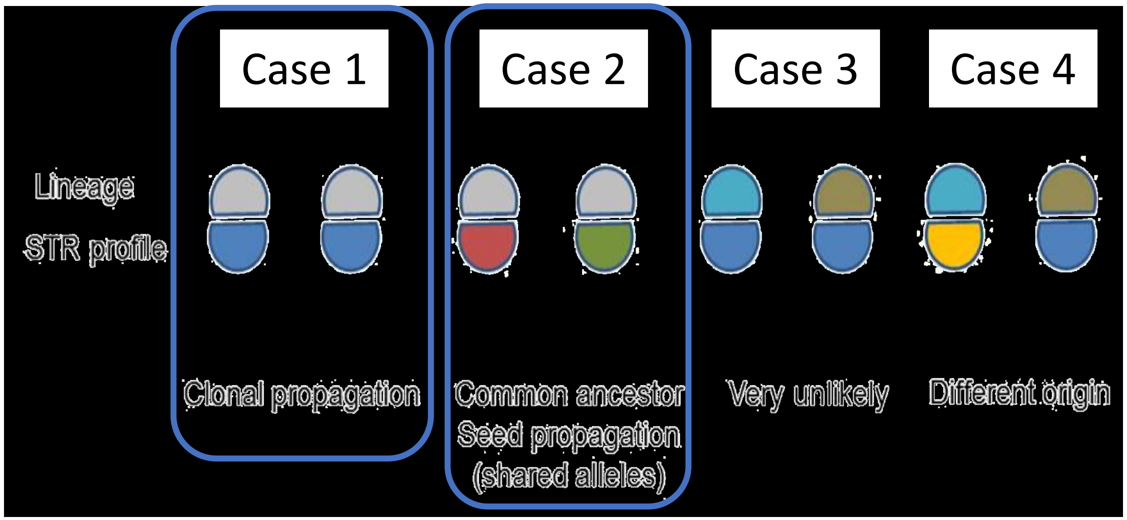
- Did not follow ISFG-SWGDAM recommendations:
  1. Avoiding the use of dinucleotides
  2. Relevant population and forensic parameters in a representative homogeneous population of *C. sativa*
  3. Real-time PCR quantitation
  4. Sequenced allelic ladders for accurate designation of alleles and inter-laboratory profile sharing

SWGDAM Scientific Working Group DNA Analysis Methods

iSFG International Society for Forensic Genetics

Purpose and goals

- Provide forensic DNA community a comprehensive analytical tool to genetically identify *C. sativa* samples:
  1. Presence of clones
  2. Association between group of samples



Materials and methods

- Sampling (4 sources)
- Extraction (DNeasy® Plant Mini Kit)
- Quantitation
- 13 STR Multiplex
- Validation Studies
- Phylogenetic Analysis

STR genotyping

Extraction
DNeasy® Plant Mini Kit

Quantitation (SYBR™ Green)
StepOne™ Real Time PCR Instrument

PCR (Type-It™ Microsatellite Kit)
Eppendorf® MasterCycler® Gradient

Capillary Electrophoresis
ABI 3500 Genetic Analyzer

Data Analysis
Genemapper® v 4.1

DNA quantitation

- Development of real-time qPCR system using SYBR™ Green with marijuana specific primers (ANUCS304) (Howard et al. (2008))
- Validation studies according to SWGDAM guidelines 9.4 and 9.5
  - Reproducibility and precision
  - Sensitivity
  - Species specificity

13 STR multiplex

Primer Selection and Optimization

Allele Sequencing and Ladder Design

Validation Studies

STR Genotyping

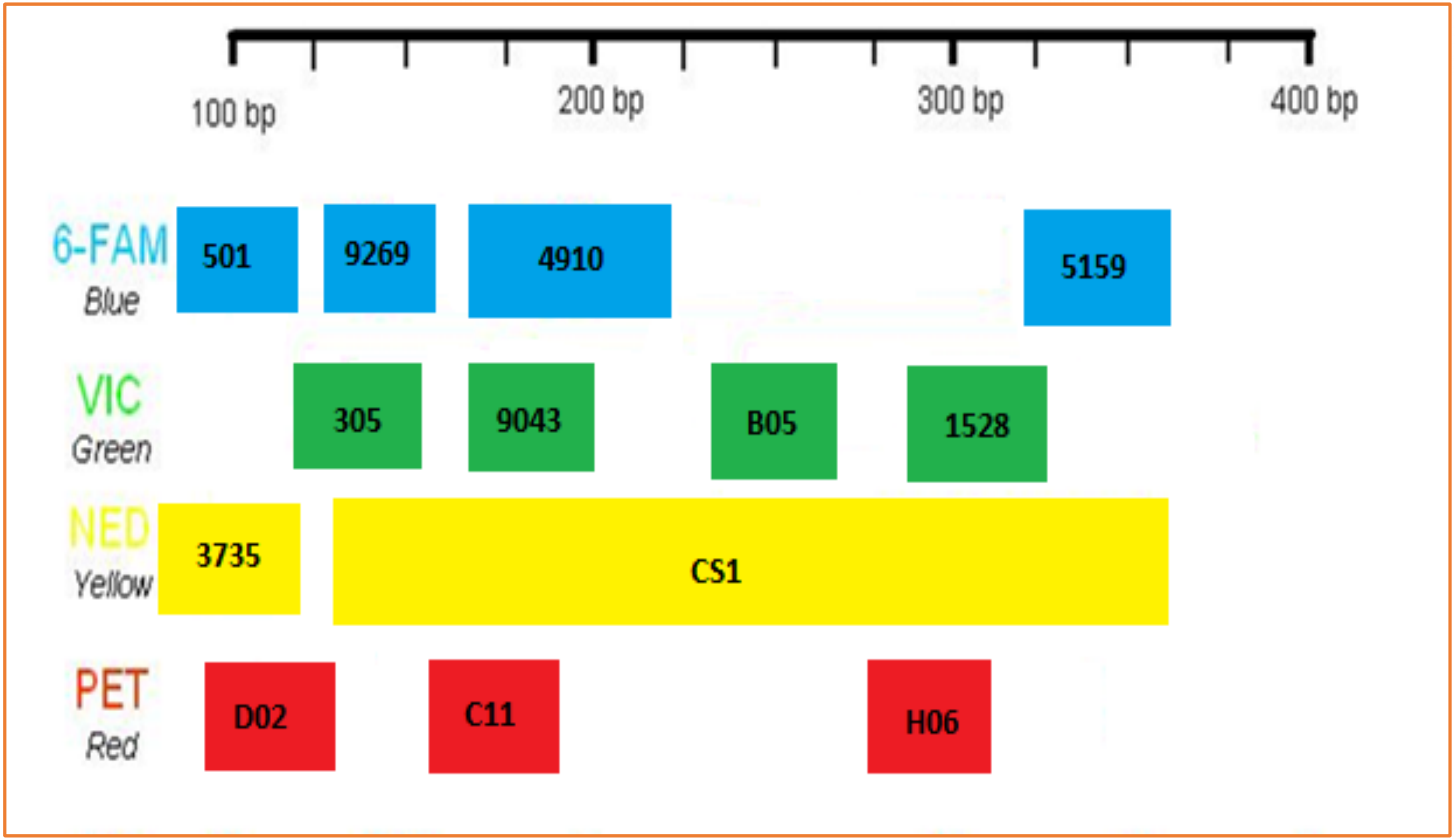
Primer selection

Type	Marker
Dinucleotide	
Trinucleotide	ANUCS305, B05, D02, C11, H06
Tetranucleotide	**9269, 4910, 5159, 9043, 1528, 3735**
Pentanucleotide	ANUCS501
Hexanucleotide	CS1

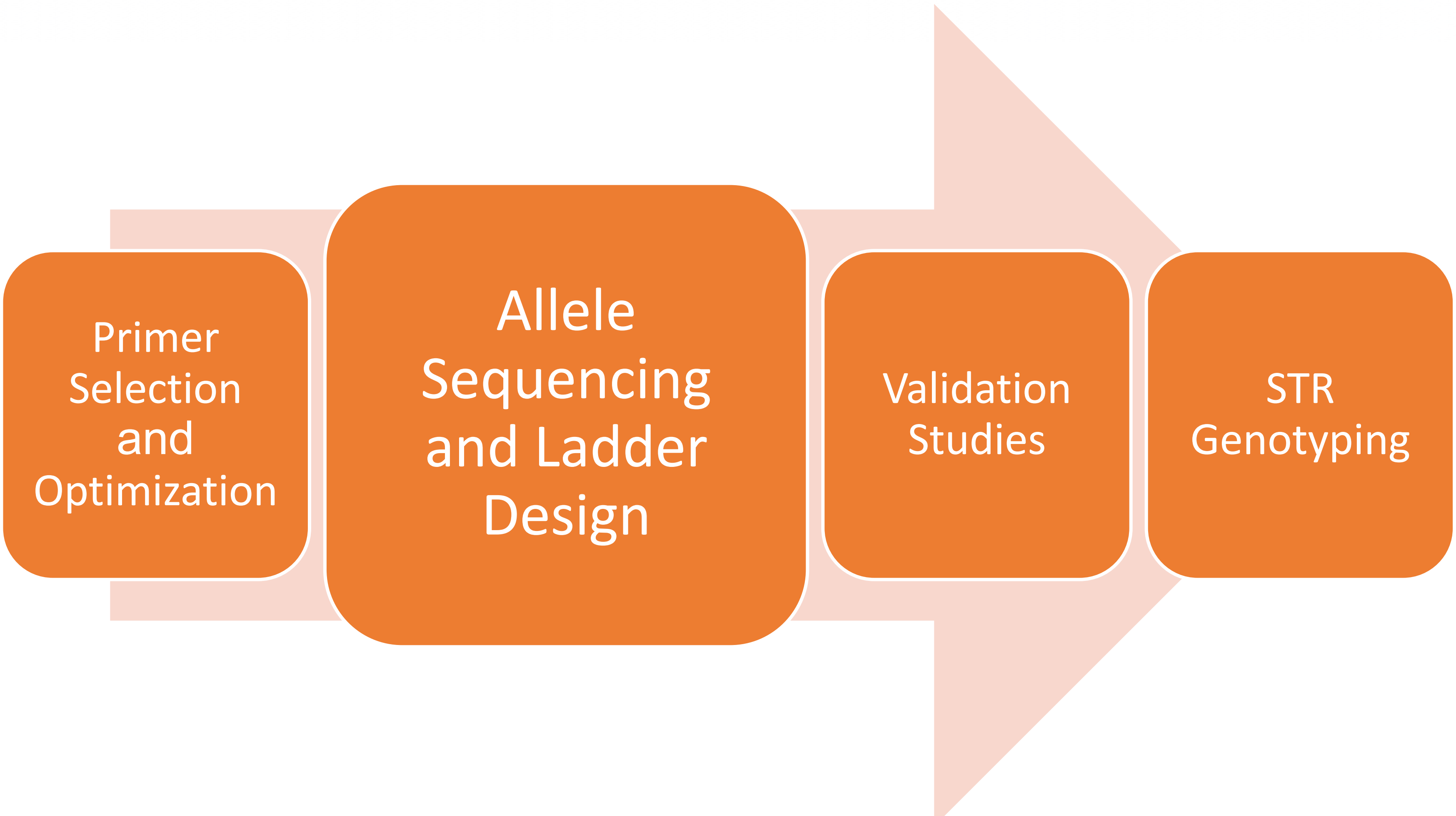
Multiplex optimization

- Multiplex Manager Software v 1.2:
  - Evaluate primer-primer interactions
  - Optimal loci layout
- Annealing temperature determination:
  - Individual loci: 65 ˚C – 55 ˚C
  - Multiplex annealing temperature: 57 ˚C
- Primer titration and cycle number:
  - Type-It™ Microsatellite PCR Kit (Qiagen)

Final 13-plex



13 STR multiplex



Sequencing/allelic ladder

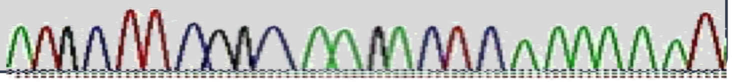
Two to eight alleles per marker

Big Dye Direct® Cycle Sequencing Kit

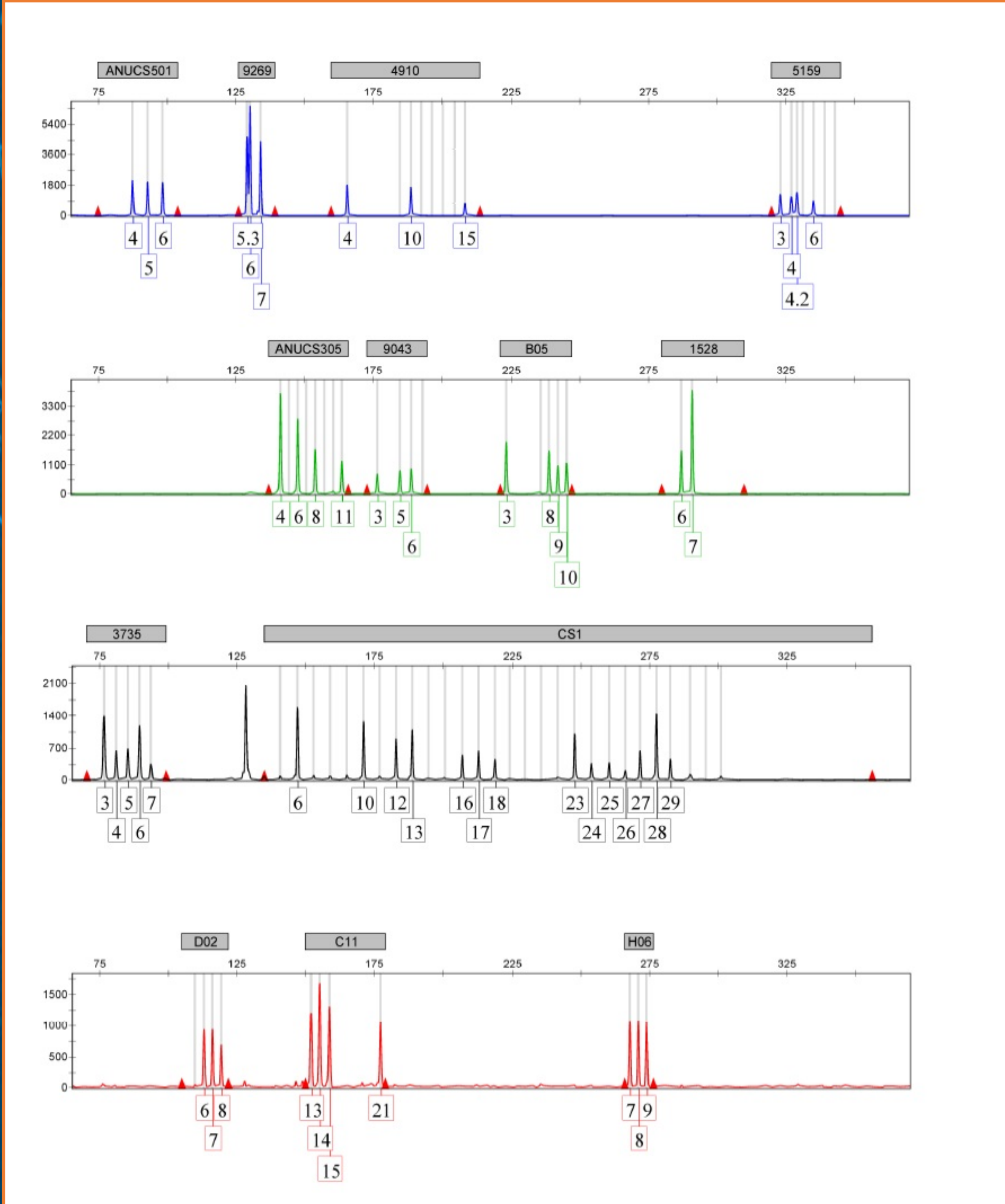
Centri-Sep™ purification columns

Geneious Pro Software

Allelic ladder design



Allelic Ladder



13 STR multiplex

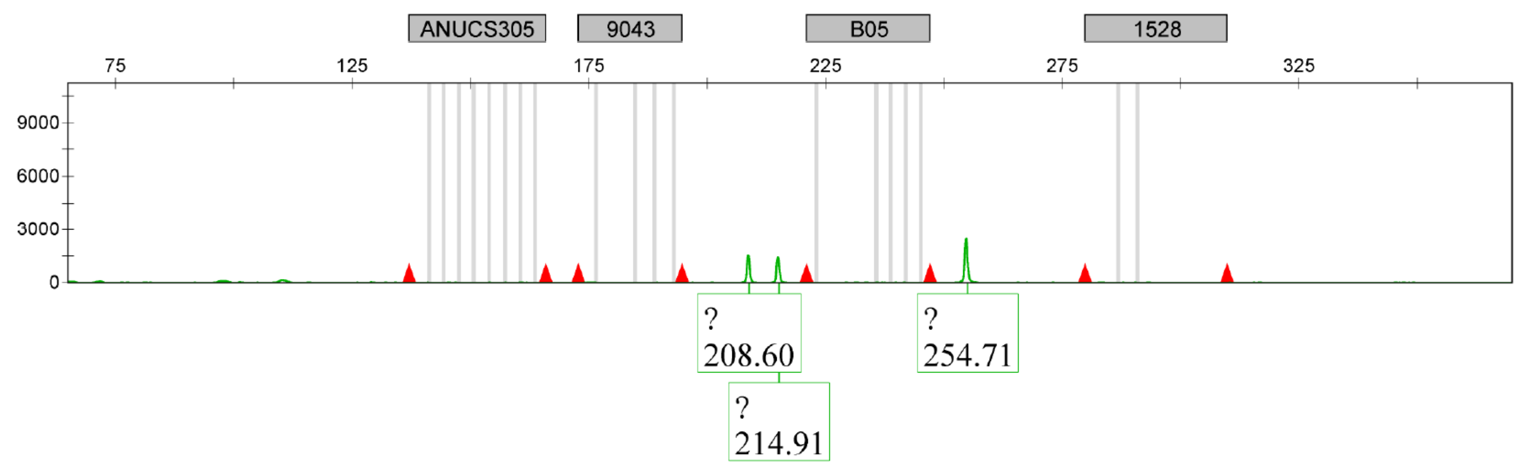
- Primer Selection and Titration
- Allele Sequencing and Ladder Design
- **Validation Studies**
- STR Genotyping

Validation studies

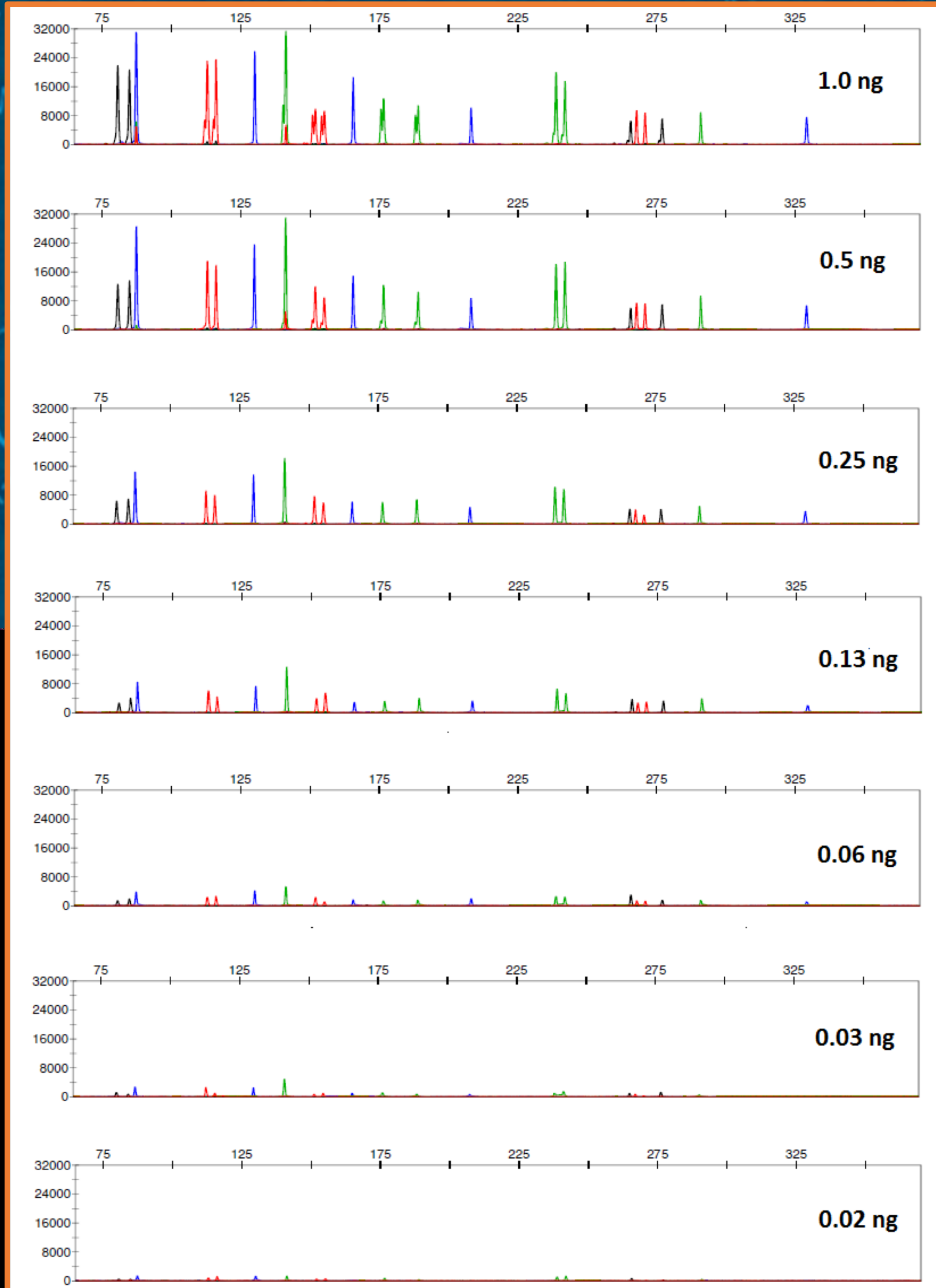
- **Developmental Validation (SWGDAM):**
  - Species specificity
  - Sensitivity/stochastic effects
  - Precision and accuracy
- **Internal Validation (SWGDAM):**
  - Stutter ratio
  - Peak height ratio
  - Inter-loci balance

Species specificity

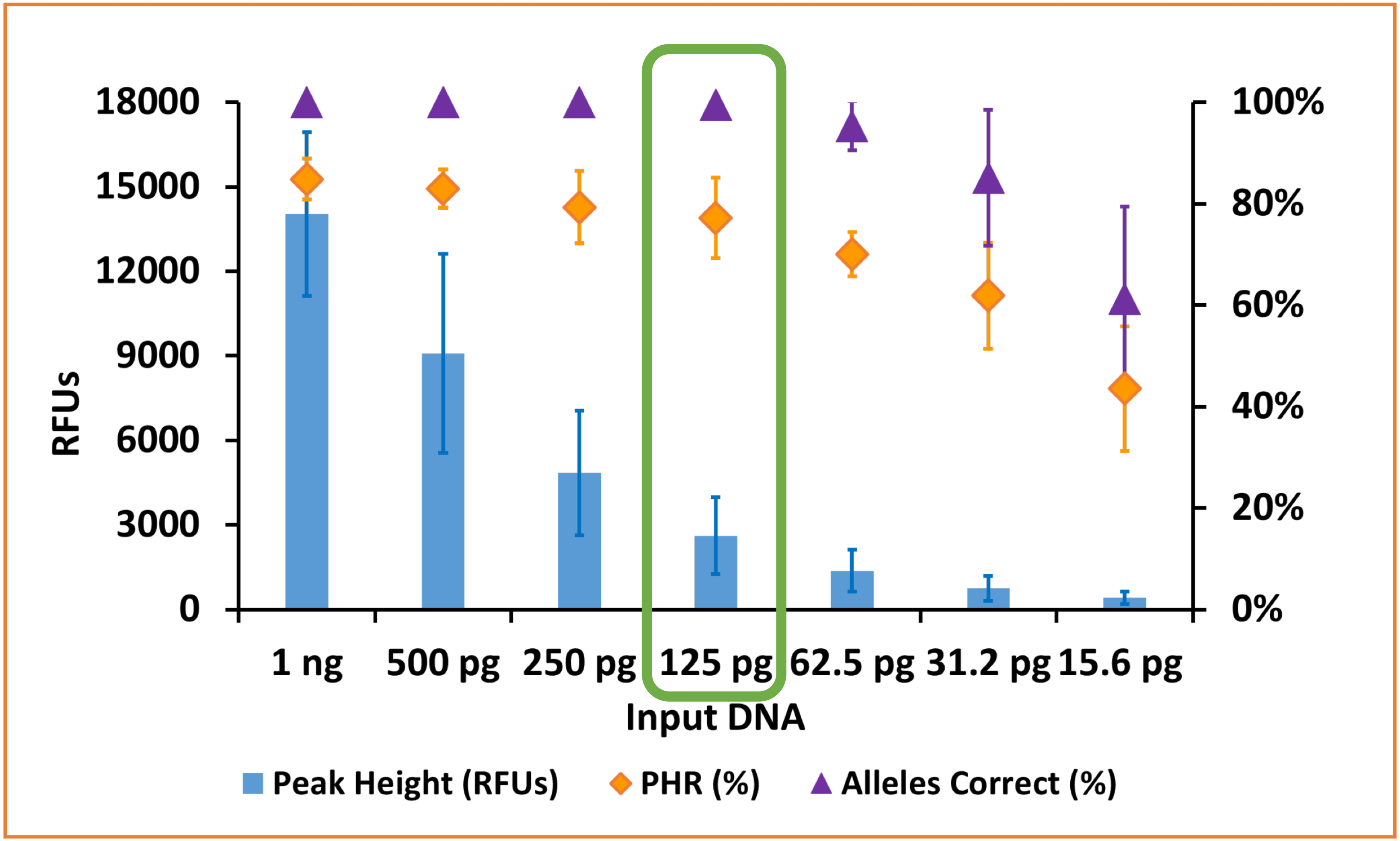
- Cross-reactivity observed in one species:
  - *Humulus lupulus* (Hops)
  - Generated non-specific peaks



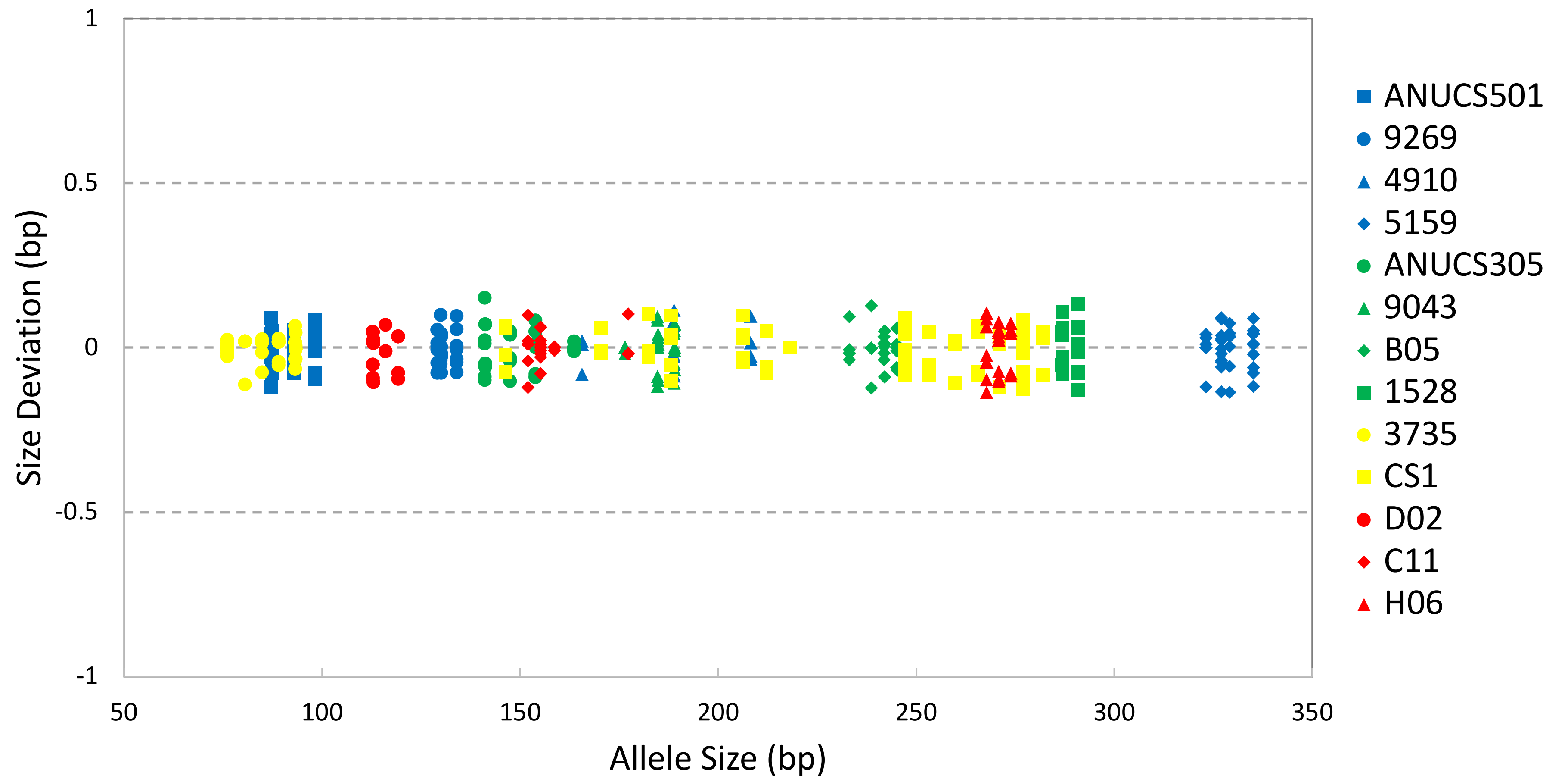
Sensitivity



Sensitivity (cont.)

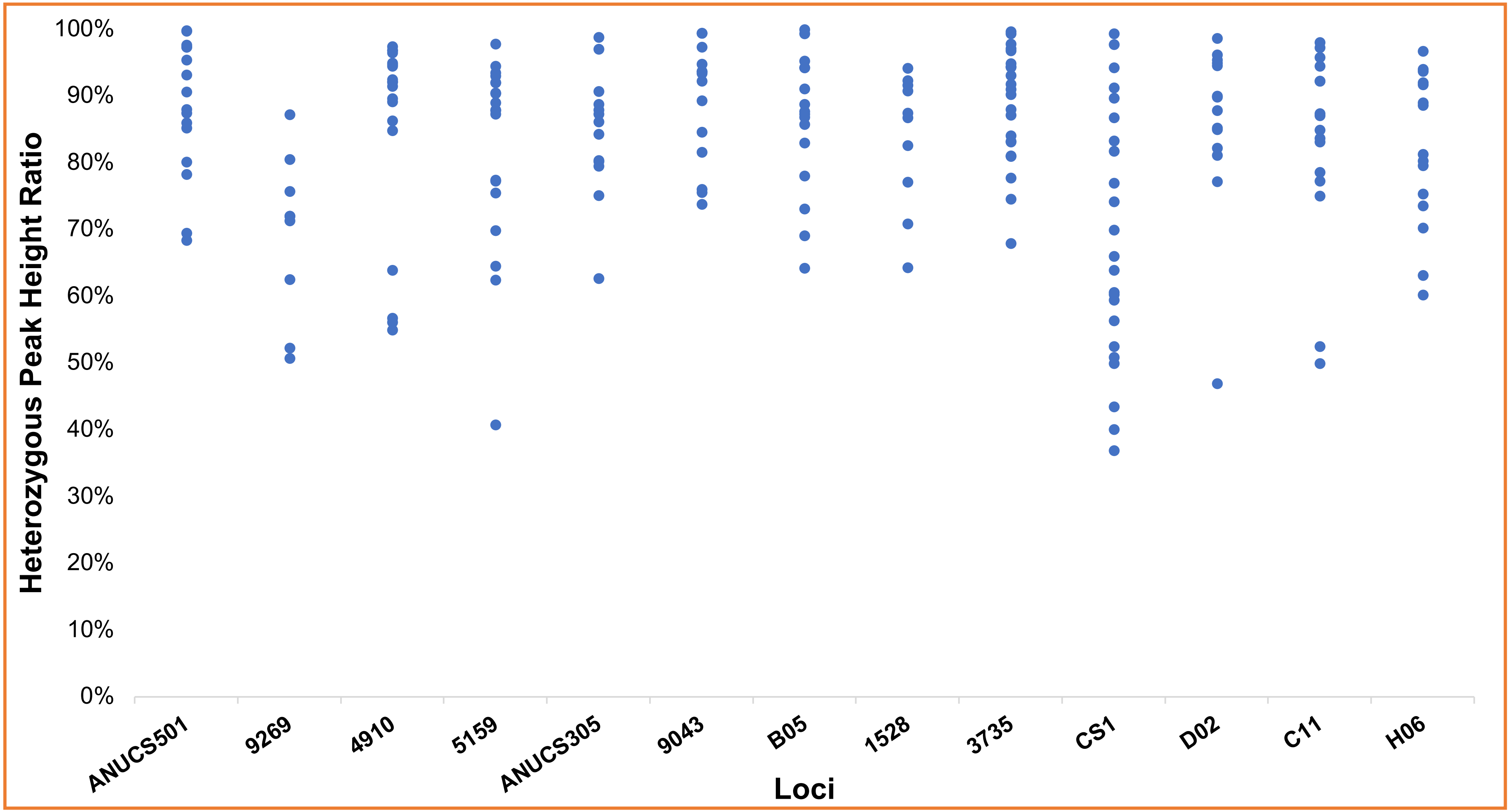


Precision & accuracy



Stutter

Peak Height Ratio



13 STR multiplex

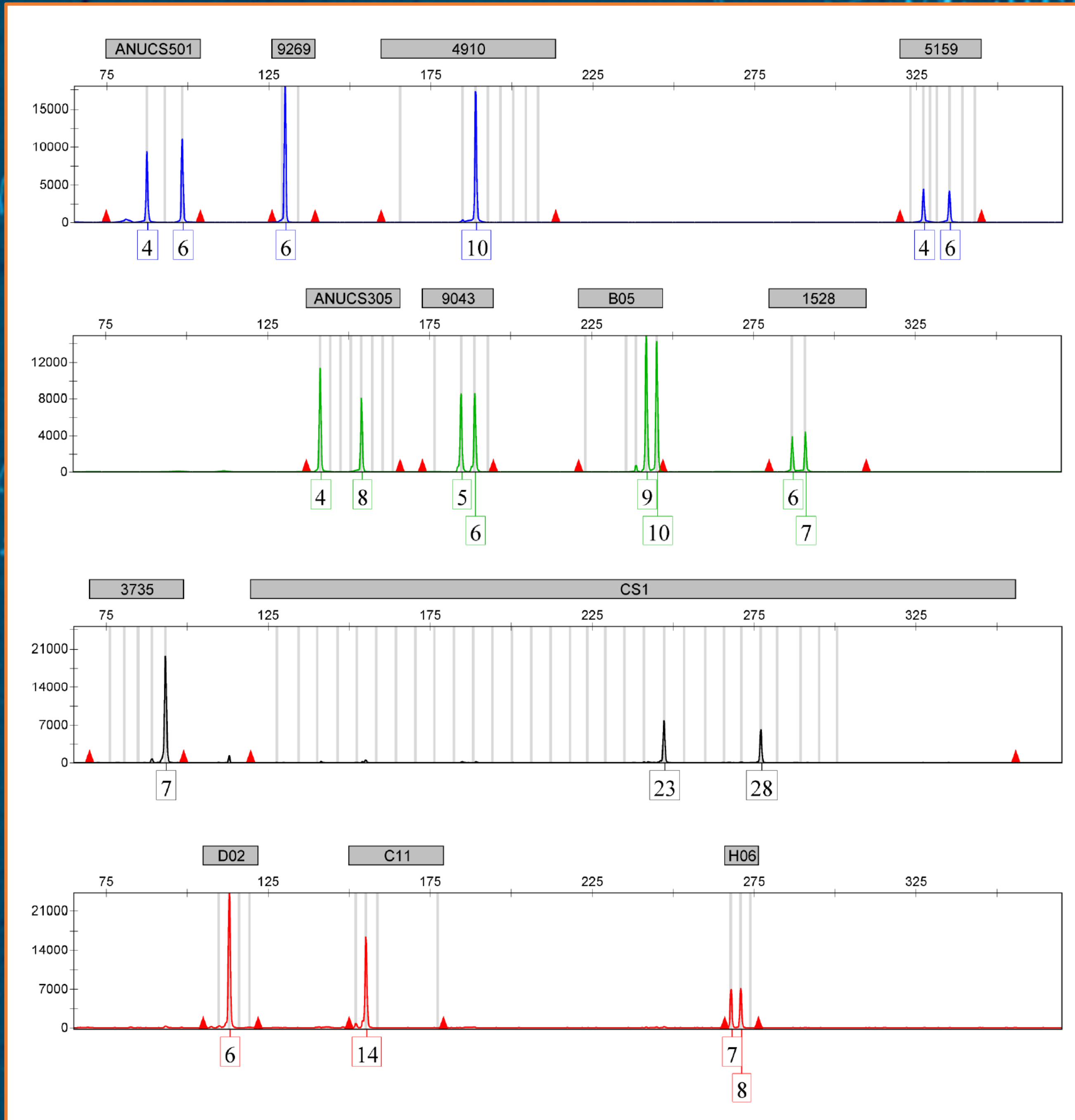
Primer Selection and Titration

Allele Sequencing and Ladder Design

Validation Studies

STR Genotyping

Positive Control



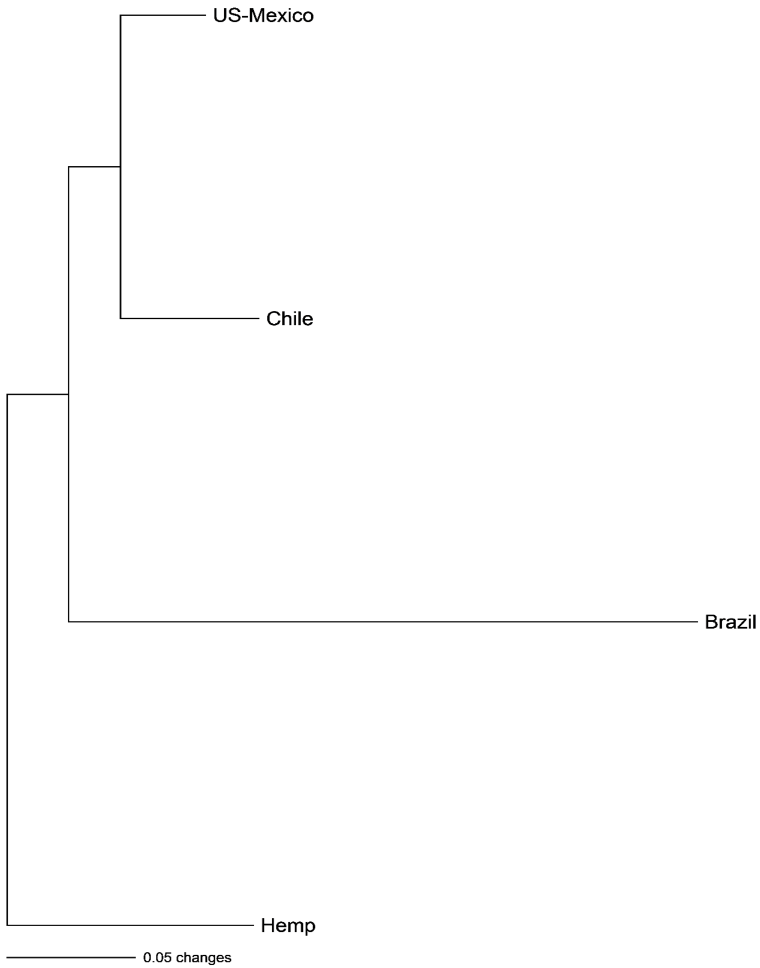
Autosomal – genotypes

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)**

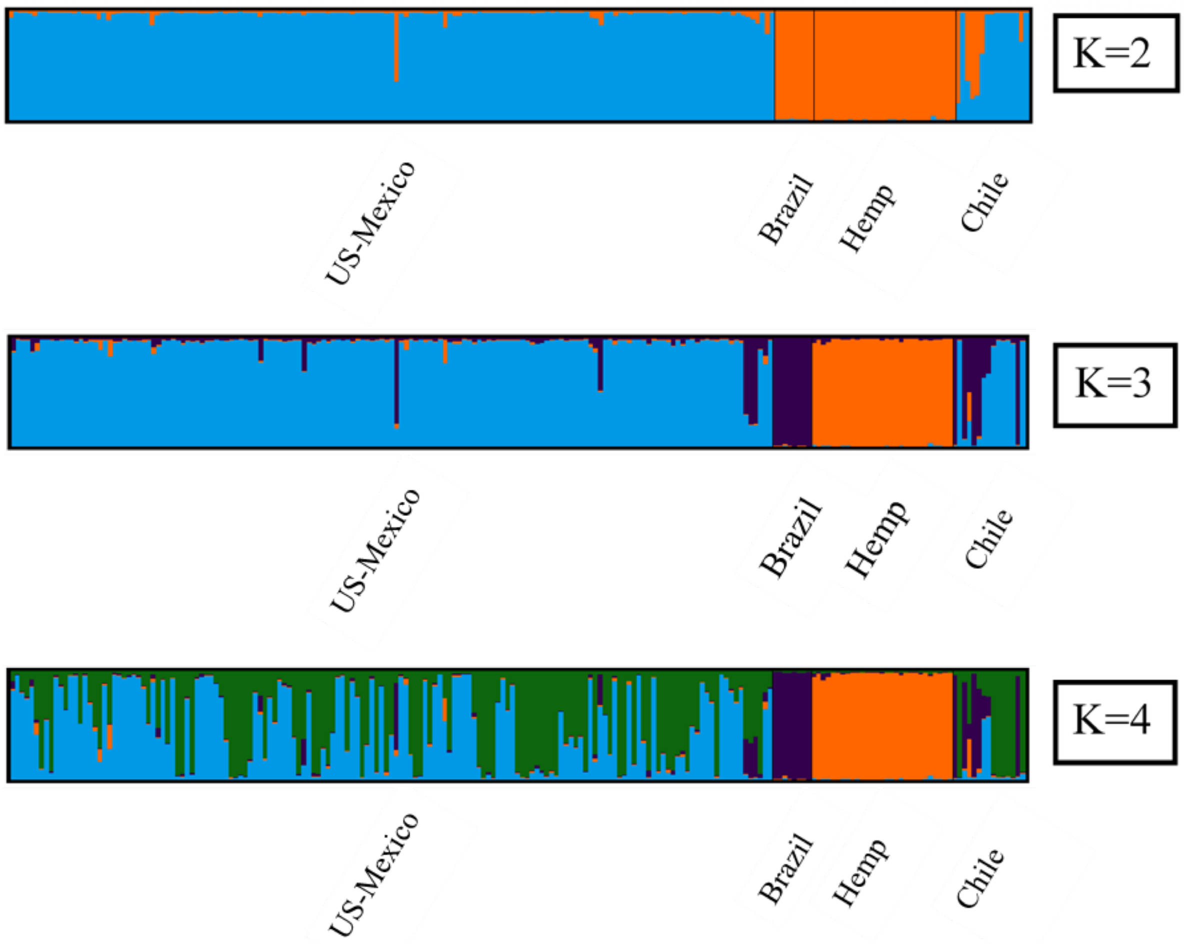
- Duplicates from same seizure or same plant
- Hypothesized that 9 out of 10 Chilean seizures clonally propagated

Autosomal – phylogenetic analysis

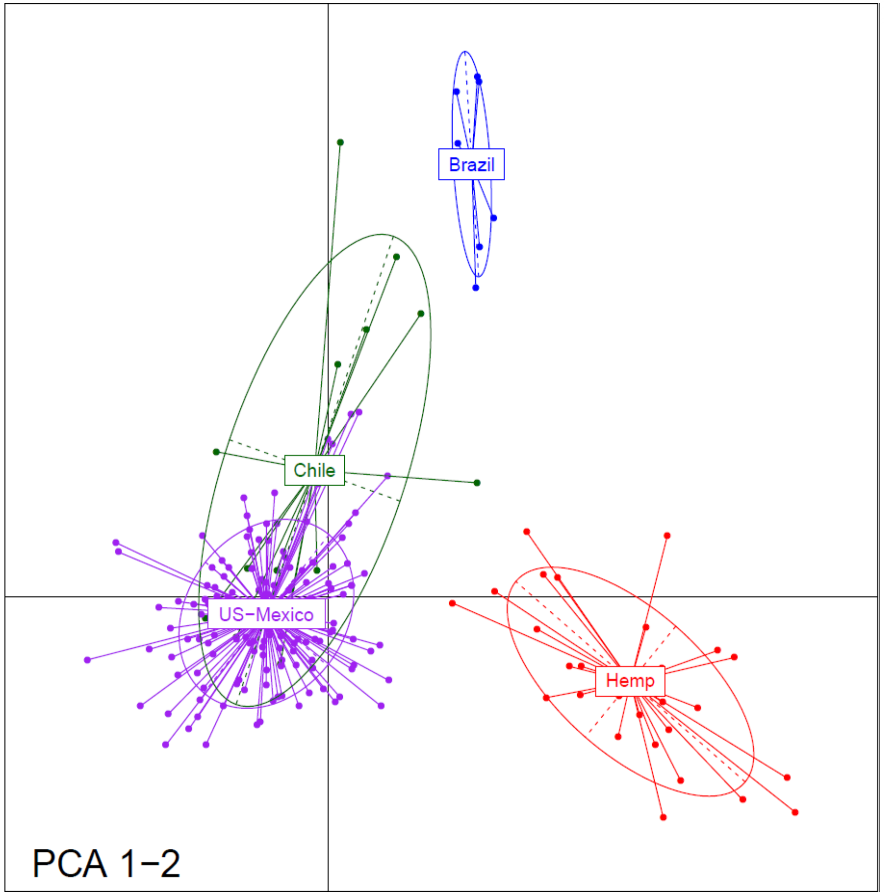
- 9 seizures (N = 157 samples) formed a reference population for US-Mexico
  - $F_{ST} = 0$



Autosomal – Bayesian clustering



Autosomal – PCA



- Genetic relatedness between drug types
- Distinction from fiber type

Statistical analysis of reference database

- Population genetic statistics and parameters of forensic interest:
  - Hardy-Weinberg Equilibrium (GDA)
  - Linkage Equilibrium (GDA)
  - Allele Frequencies (PowerStats v 1.2)
  - Random Match Probability (PowerStats v 1.2)
  - Power of Discrimination (PowerStats v 1.2)

Combined power of discrimination of the multiplex is 1 in 55 million

Potential impact

- Provide the forensic community with a genetic tool for identification of *C. sativa* samples:
  1. Authenticate legal cannabis products
  2. Link cases (as intelligence tool)
  3. Link and identify illegal growers/distributers
- Complement previously established profiling programs for intelligence purposes for organizations, such as Homeland Security/CBP and DEA

Legal Medicine 26 (2017) 33–40

Contents lists available at ScienceDirect

Legal Medicine

journal homepage: www.elsevier.com/locate/legalmed

Short Communication

Developmental and internal validation of a novel 13 loci STR multiplex method for *Cannabis sativa* DNA profiling

Rachel Houston [a], Matthew Birck [b], Sheree Hughes-Stamm [a], David Gangitano [a,*]

[a] *Department of Forensic Science, College of Criminal Justice, Sam Houston State University, 1003 Bowers Blvd., Huntsville, TX 77340-2525, United States*
[b] *New York Laboratory, U.S. Customs and Border Protection, U.S. Department of Homeland Security, 1100 Raymond Blvd, Newark, NJ 07102, United States*

International Journal of Legal Medicine (2018) 132:713–725
https://doi.org/10.1007/s00414-018-1798-4

ORIGINAL ARTICLE

Nuclear, chloroplast, and mitochondrial data of a US cannabis DNA database

Rachel Houston [1] · Matthew Birck [2] · Bobby LaRue [1] · Sheree Hughes-Stamm [1] · David Gangitano [1]

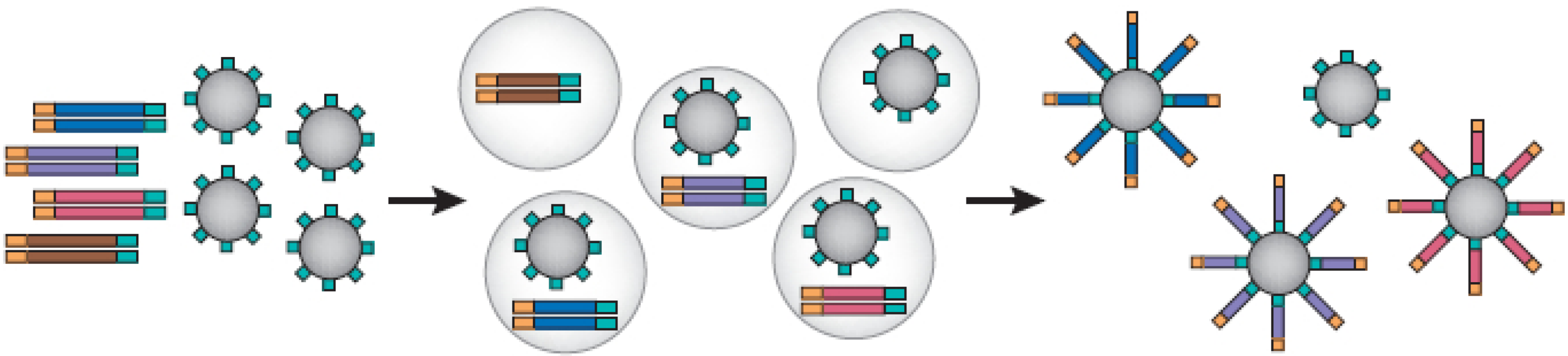
Received: 9 November 2017 / Accepted: 9 February 2018 / Published online: 20 February 2018

Massively parallel sequencing of 12 autosomal STRs in *Cannabis sativa*

SeqWeed: The Next Generation

Why massively parallel sequencing (MPS)?

- High throughput and multiplex capability
- Greater power of discrimination
- Degraded DNA
- Mixture interpretation



Source	N
US-Mexico	8
Brazil	2
Chile	3
Hemp	3

N = 16 samples

Extract & Quantify DNA

Targeted Amplification

Custom PCR assay with Type-It™ Microsatellite Kit (Qiagen)

Library Preparation

Ion Plus Fragment Kit

Clonal Amplification/Templating

Sequencing

Data Analysis

STR panel

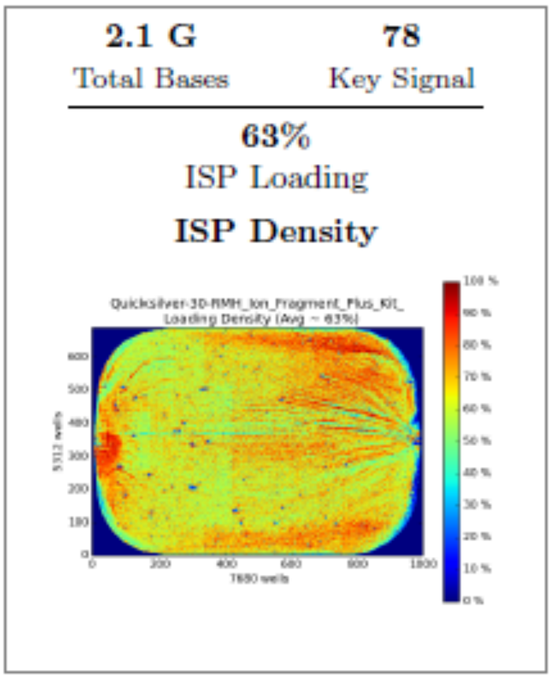
- 12plex STR panel:
  - ANUCS501
  - 9269
  - 4910
  - 5159
  - ANUCS305
  - 9043
  - B05-CANN1
  - 1528
  - 3735
  - D02-CANN1
  - C11-CANN1
  - H06-CANN2
- Analyzed with:
  - STRait Razor v3
  - Excel workbook
- Compared data to genotypes generated on 3500 Genetic Analyzer
- Evaluation of sequence variants within alleles

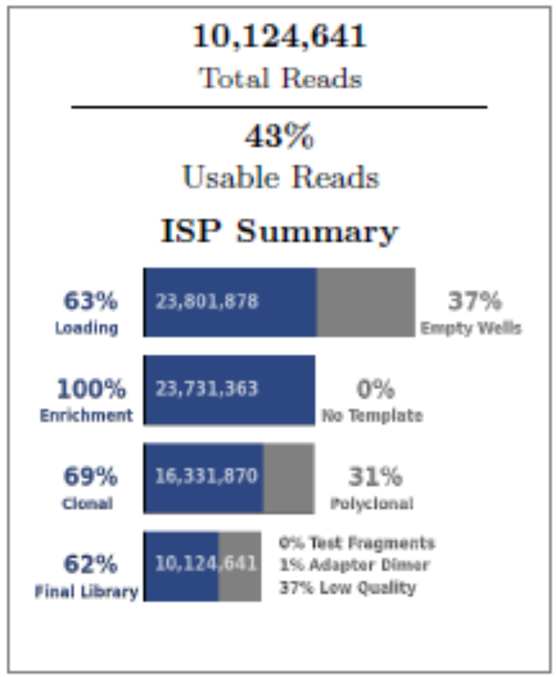
Data sorting and analysis

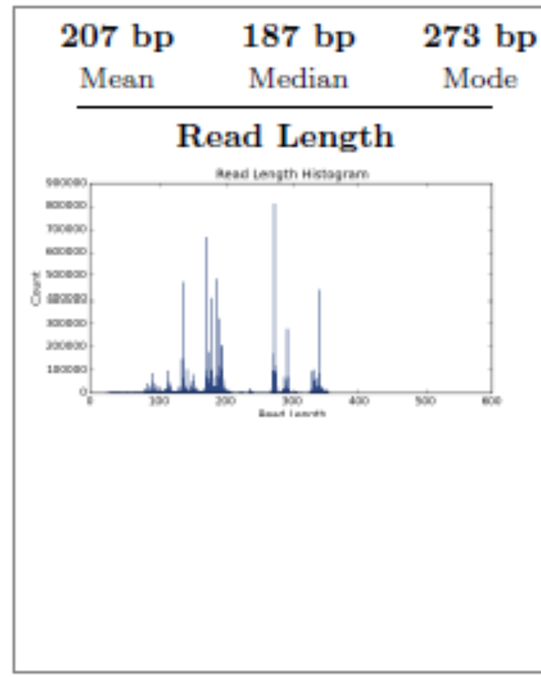
- Sequencing data generated on Torrent Suite Software (v.5.2.2)
- FASTQ files exported to bioinformatics tool (STRait Razor v3)
  - Customizable
  - Extract and parse STR data
- STRait Razor Excel workbook
  - Collate and visualize data

Sequencing data

Torrent Suite Software:

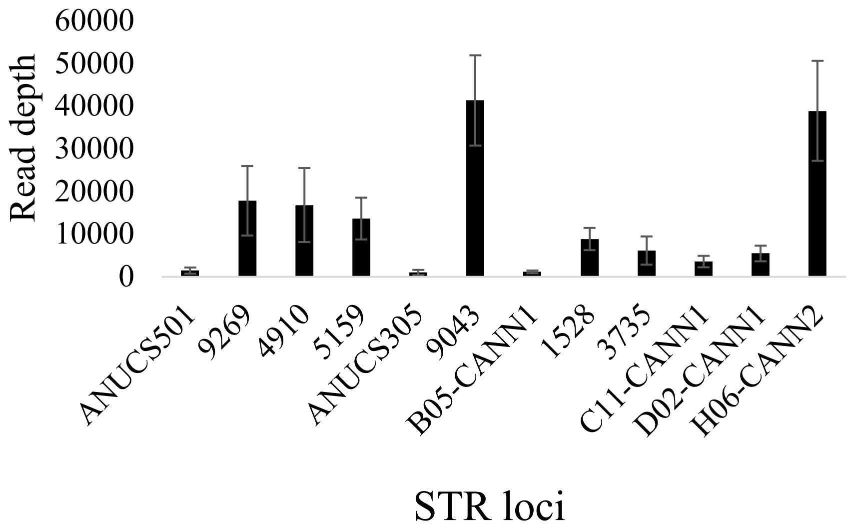






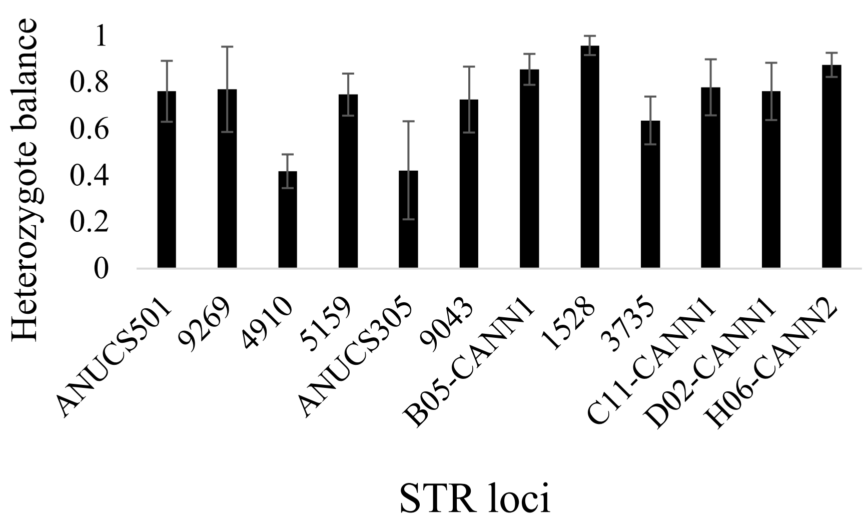
Read depth and heterozygote balance

Read Depth



- Average: 13,000x
- Range: 1,001 – 41,378x

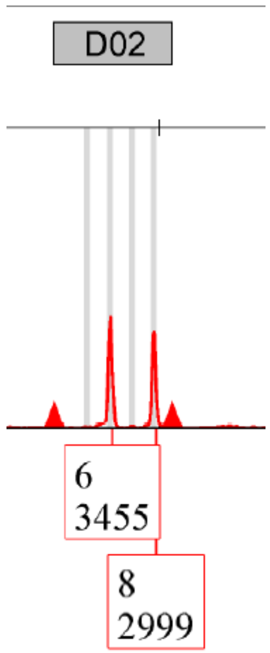
Heterozygote balance

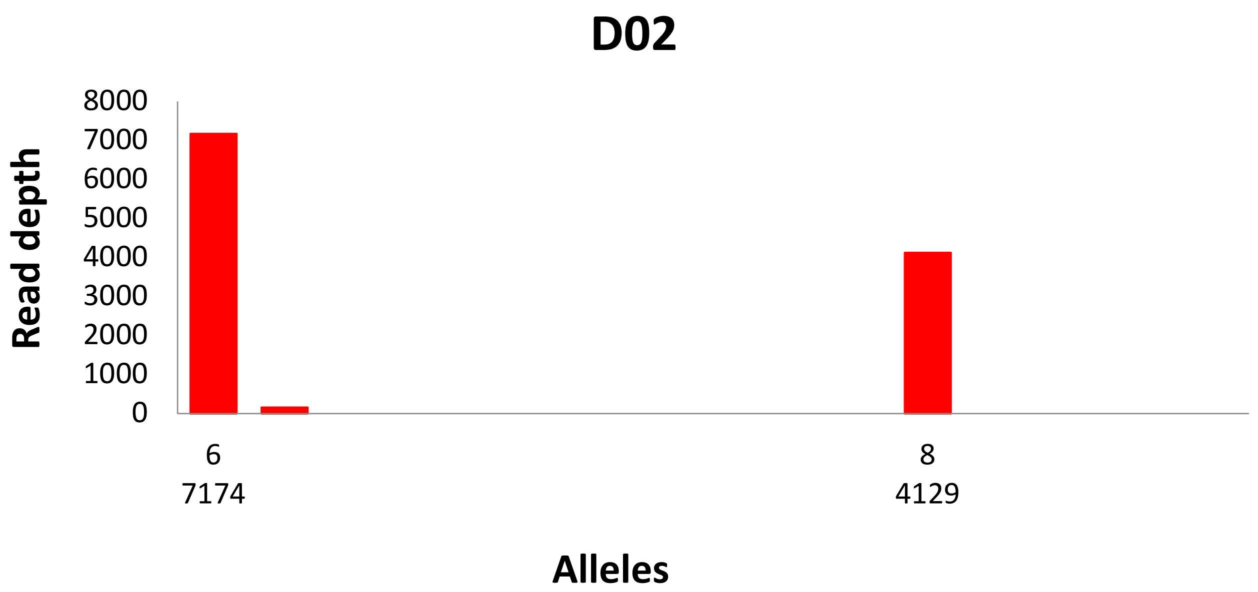


- Average: 0.73 ± 0.16

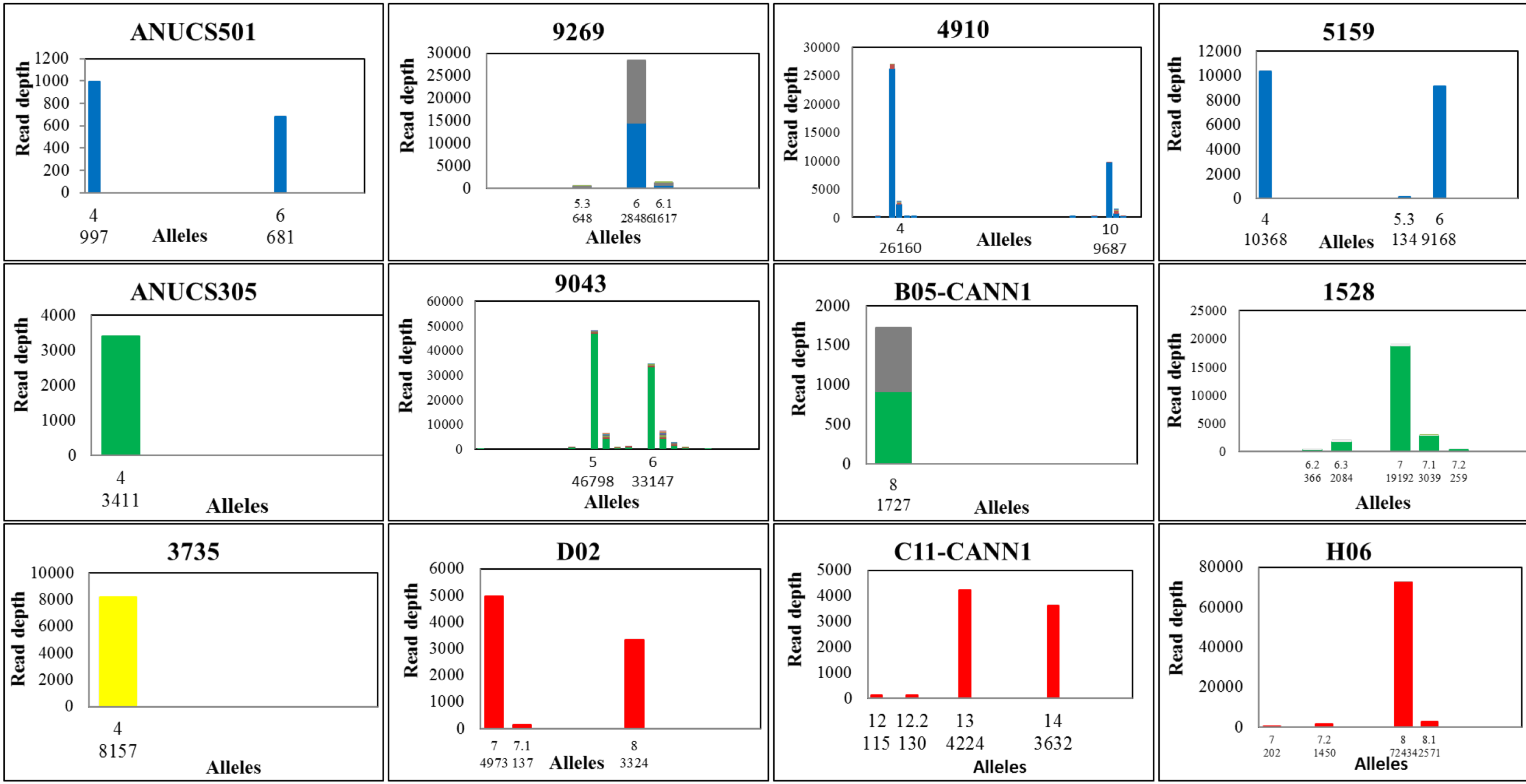
STRait Razor analysis

- Comparable methods of visualization for CE and MPS platforms





Example histogram



STR results

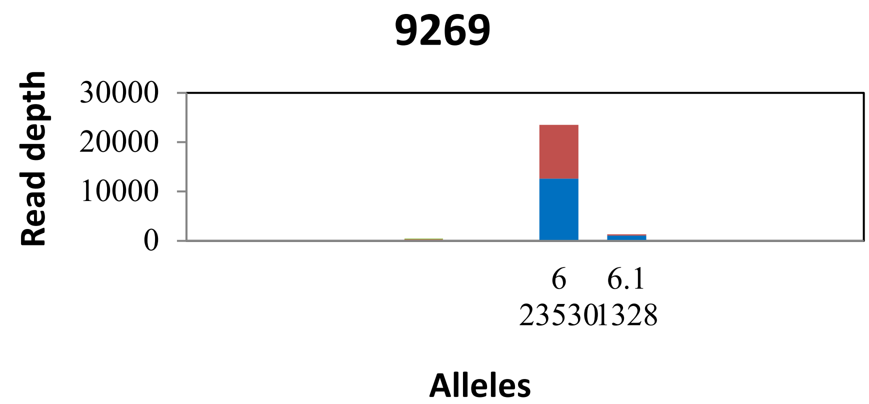
Sample	A501	9269	4910	5159	A305	9043	B05	1528	3735	D02	C11	H06
1	4,4	6,6	4,10	4.2,6	4,6	6,6	8,8	6,7	4,4	6,8	13,14	8,8
2	4,6	6,6	4,10	6,6	4,6	5,5	8,9	6,7	5,7	7,7	13,13	7,8
3	6,6	5.3,6	4,10	4.2,6	4,8	5,6	8,9	6,6	6,7	6,7	13,13	8,9
4	4,4	6,6	10,10	6,6	4,8	3,5	8,9	7,7	3,6	6,7	14,14	7,8
5	4,6	6,6	4,10	4,6	4,4	5,6	8,8	7,7	4,4	7,8	13,14	8,8
6	4,4	6,6	4,4	3,3	6,8	3,6	8,8	7,7	3,5	6,6	13,13	8,9
7	4,4	5.3,5.3	5,5	6,6	4,4	5,5	7,8	7,7	6,6	8,8	14,14	8,8
8	2,4	4,6	5.3,5.3	6,7	8,8	5,6	8,9	7,7	8,8	6,6	14,14	8,8
9	4,6	6,6	4,10	4,4.2	4,8	3,3	8,8	7,7	4,7	6,6	13,14	7,8
10	6,6	6,6	4,4	4,4.2	4,8	3,6	8,8	7,7	4,7	6,7	14,14	8,8
11	4,6	6,6	10,10	6,6	8,8	5,6	8,9	7,7	5,6	7,7	13,14	7,8
12	4,6	6,6	4,10	4.2,6	4,8	3,3	8,8	6,7	3,5	6,7	14,15	8,8
13	4,6	6,6	4,10	4.2,6	4,8	3,3	8,8	6,7	3,5	6,7	14,15	8,8
14	2,4	6,6	10,10	6,6	7,9	6,6	8,9	7,7	4,6	5,6	16,16	8,8
15	2,4	6,6	10,10	6,7	6,9	5,5	9,9	7,7	4,6	6,6	16,16	8,8
16	4,4	5.3,5.3	5,5	6,6	4,4	5,5	8,8	7,7	6,6	8,8	14,14	8,8

All genotypes in concordance with CE data

Isoalleles detected

Sequence variants

- 8 loci where nominal allele differentiated by sequence
  - 9269
  - 5159
  - ANUCS305
  - 9043
  - B05-CANN1
  - 1528
  - D02-CANN1
  - C11-CANN1



Nominal Allele	String sequence	Read depth
6	CTTTCTATCAGTATCTATAAAATATTTACAAGA AAAGAGCATT [ATCA] $[ATAA]_5$	12580
6	CTTTCTATCAGTATCTATAAAATATTTACAAGA AAAGAGCATT $[ATAA]_6$	10950

Conclusions

- Successful proof of concept
- Identified sequence variants (isoalleles)
- Illustrates a workflow for integrating any PCR multiplex into MPS workflow
  - Library prep
  - Bioinformatics
- Future research:
  - Expand number of loci
  - Gao et al. identified 56 polymorphic STRs

Acknowledgements

- NIJ Award Number #2015-R2-CX-0030
- Sam Houston State University, Department of Forensic Science
- Dr. David Gangitano
- Dr. Sheree Hughes-Stamm
- Team DNA!

Questions

Rachel Houston

rmh034@shsu.edu