EVALUATION OF A 13-LOCI STR MULTIPLEX SYSTEM FOR *CANNABIS SATIVA* GENETIC IDENTIFICATION

Rachel Houston, B.S.

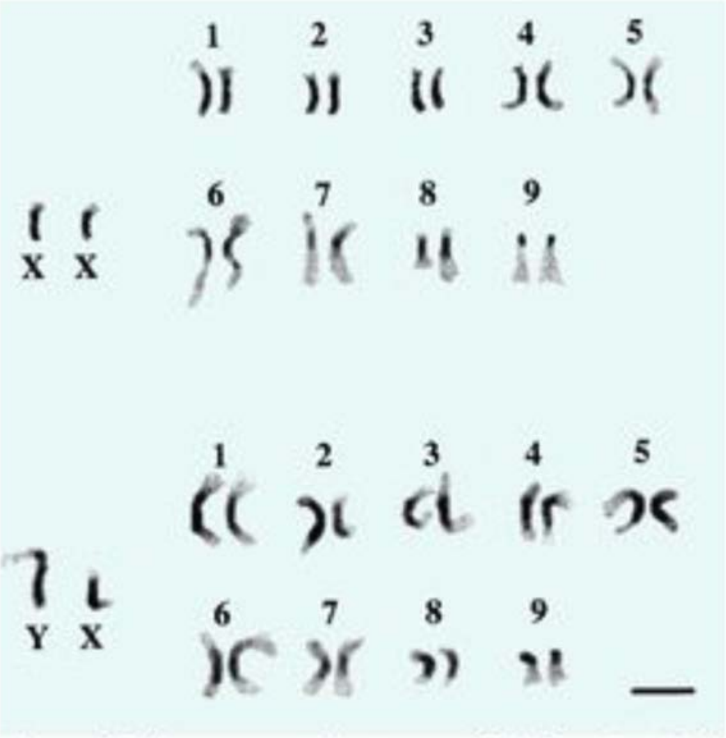
AFDAA Summer 2015

Outline

- Introduction
- Previous Work
- Multiplex Development
- Population Database
- Conclusions
- Potential Impact
- Acknowledgements

Marijuana Background

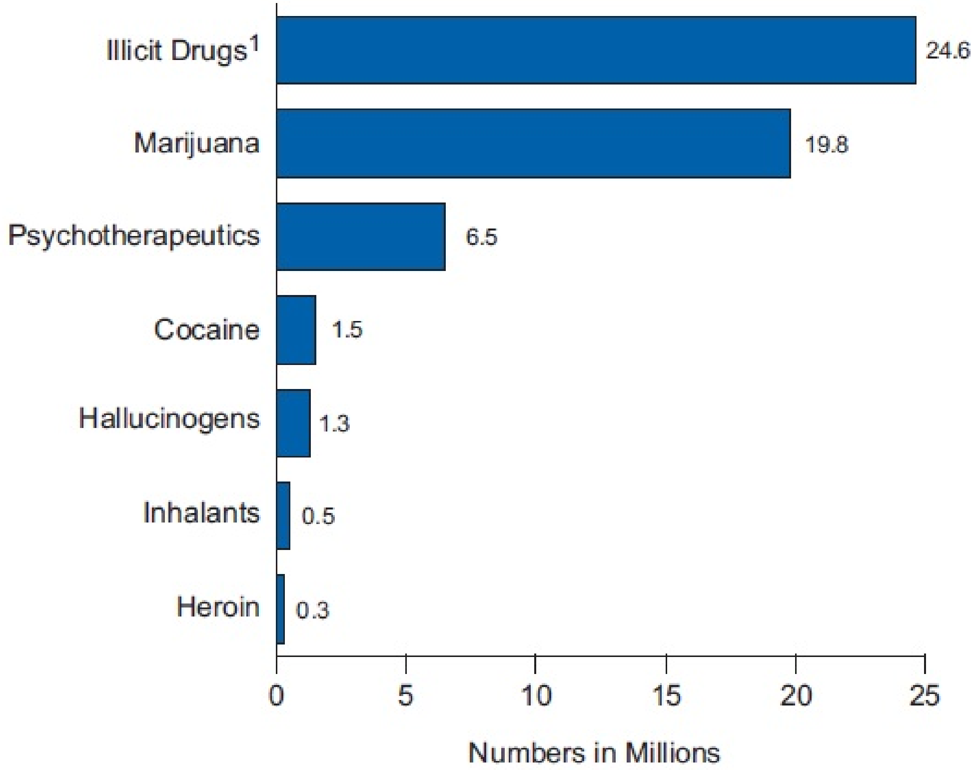
- Plant material from *Cannabis sativa*
- Family: Cannabaceae
- Genus: *Cannabis*
- Species: *Cannabis sativa*
  - Strains: sativa, indica, ruderalis
- Diploid genome (2n = 20)
  - 9 sets of autosomes
  - Pair of sex chromosomes



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

Significance

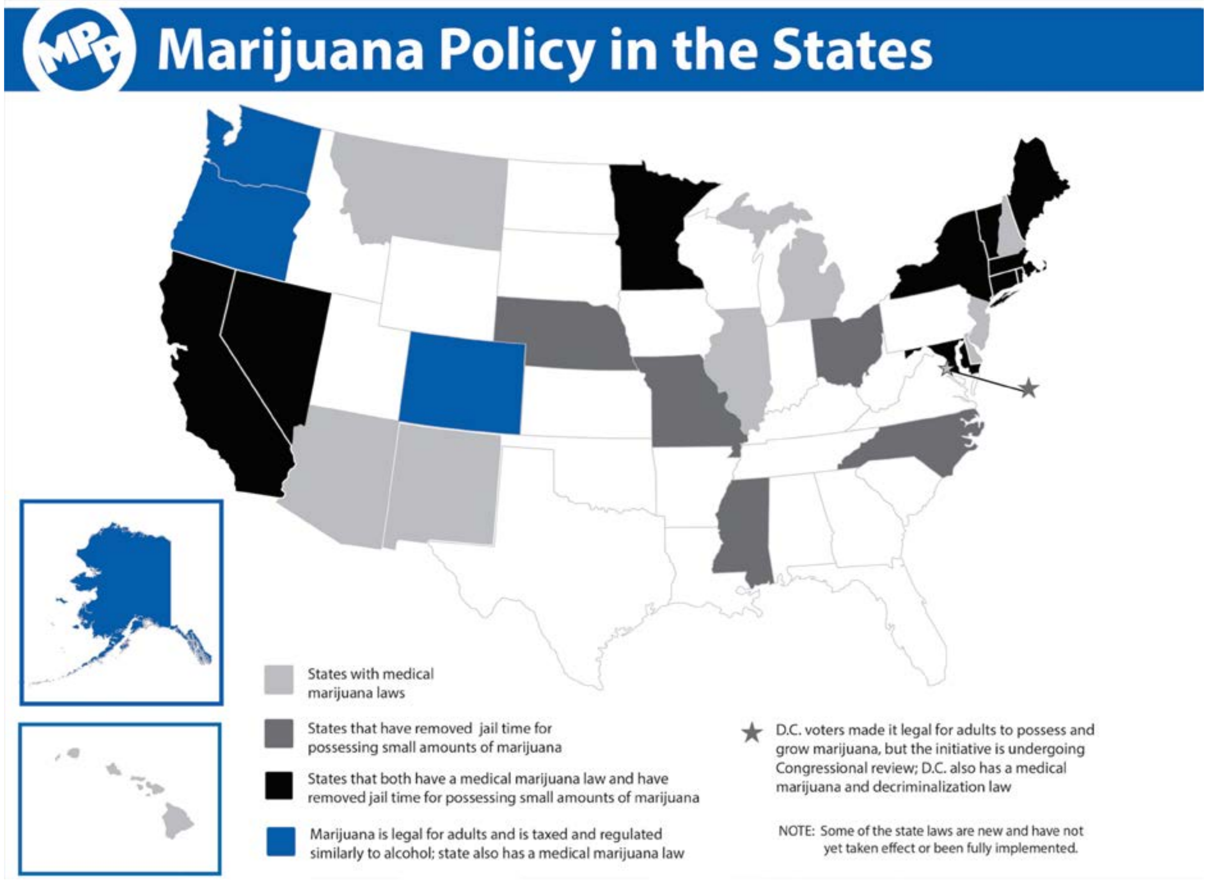
- Marijuana is the most commonly used illicit drug in United States



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

Marijuana Legalization

- Legalized for recreational use in:
    - Colorado
    - Washington
    - Alaska
    - Oregon
    - District of Columbia



Marijuana Policy in the States

▨ States with medical marijuana laws

▨ States that have removed jail time for possessing small amounts of marijuana

■ States that both have a medical marijuana law and have removed jail time for possessing small amounts of marijuana

🟦 Marijuana is legal for adults and is taxed and regulated similarly to alcohol; state also has a medical marijuana law

★ D.C. voters made it legal for adults to possess and grow marijuana, but the initiative is undergoing Congressional review; D.C. also has a medical marijuana and decriminalization law

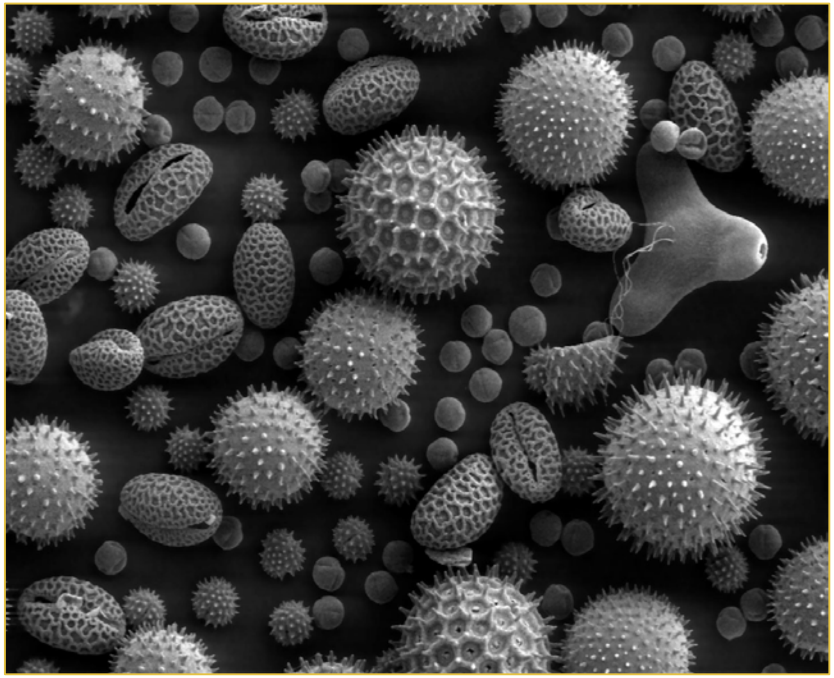
NOTE: Some of the state laws are new and have not yet taken effect or been fully implemented.

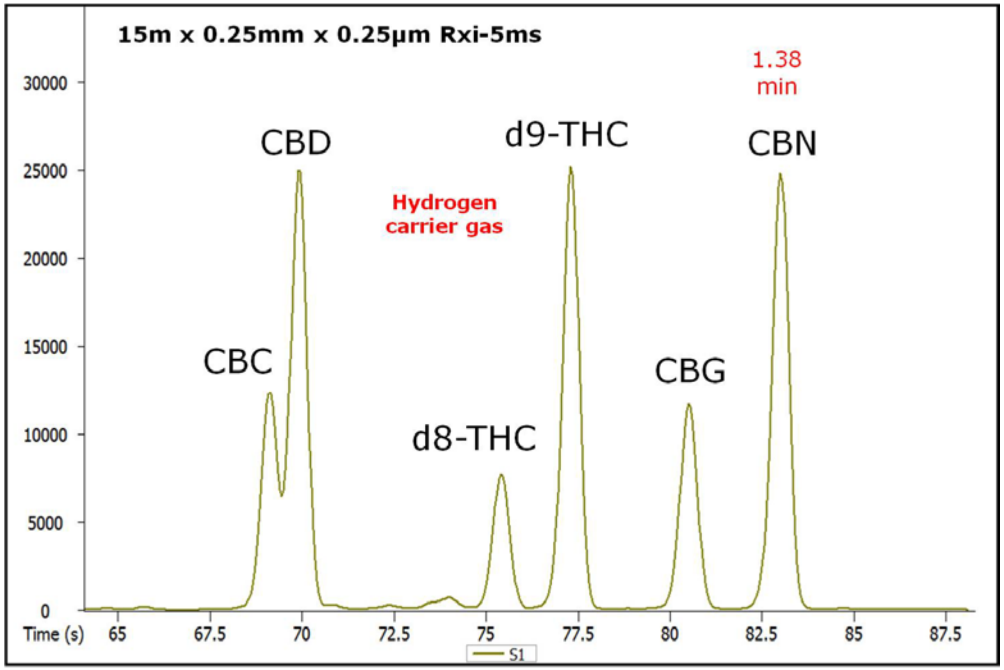
Illegal Traffic at the US-Mexico border

- Active illegal traffic at the US border
- Serious concern for US Federal government

Individualization Methods

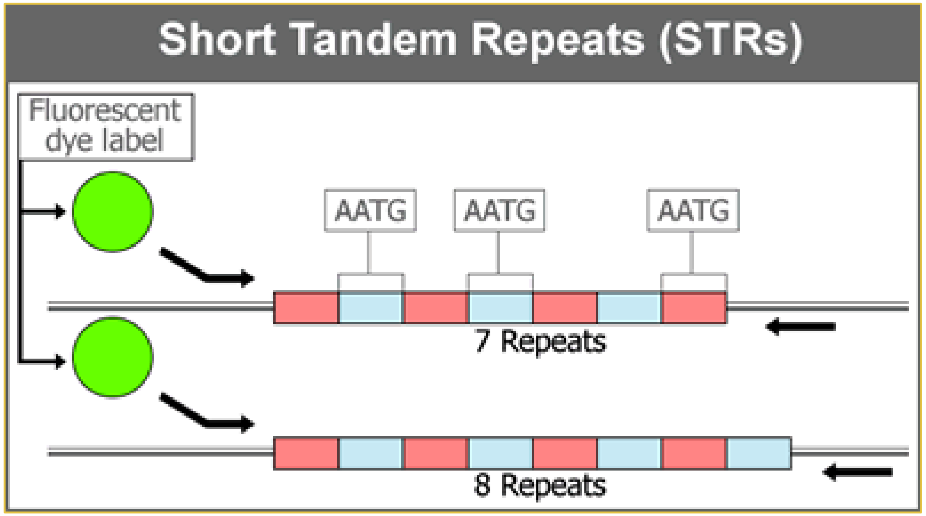
- Palynology
- Chemical profiling (GC-MS)
- Isotope Ratio Mass Spectrometry (IRMS)
    - Carbon and Nitrogen





DNA Based Individualization

- Chloroplast DNA (CpDNA)
- Mitochondrial DNA (MtDNA)
- Short Tandem Repeat (STR)
  - Gold standard in human identification



Previous STR Research

- STR polymorphic markers first described:
  - Gilmore and Peakall (2003)
  - Alghanim and Almirall (2003)
- Australia
  - Marijuana DNA STR multiplex and database (Howard *et al.*)
- United States
  - Attempted STR database (Mendoza *et al.*)
  - CS1 marker study (Coyle *et al.)*
- Germany
  - 15 loci - STR tool (Köhnemann *et al.*)
  - Proposed new tetranucleotide markers (Köhnemann *et al.*)

Limitation of Previous Research

- Did not follow ISFG-SWGDAM Recommendations:
  1. Use of sequenced allelic ladders for accurate designation of alleles and inter-laboratory STR profile sharing
  2. Relevant population and forensic parameters studied in a representative homogeneous (low *Fst*) population of *C. sativa* for random match probability estimations or verification of genetic relatedness
  3. Real-time PCR quantitation

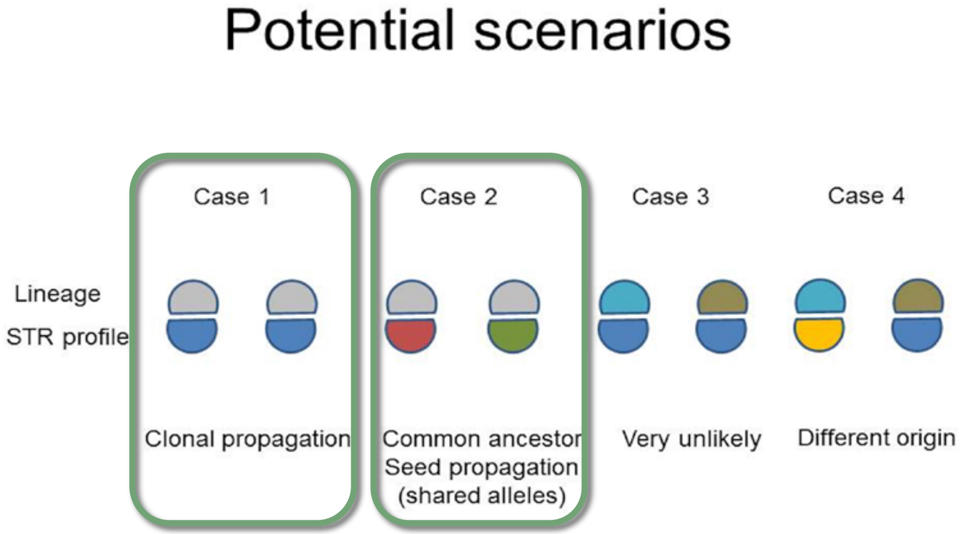
NEW!

SCIENTIFIC WORKING GROUP
SWGDAM
DNA ANALYSIS METHODS

ISFG International Society for Forensic Genetics

Purpose and Goals

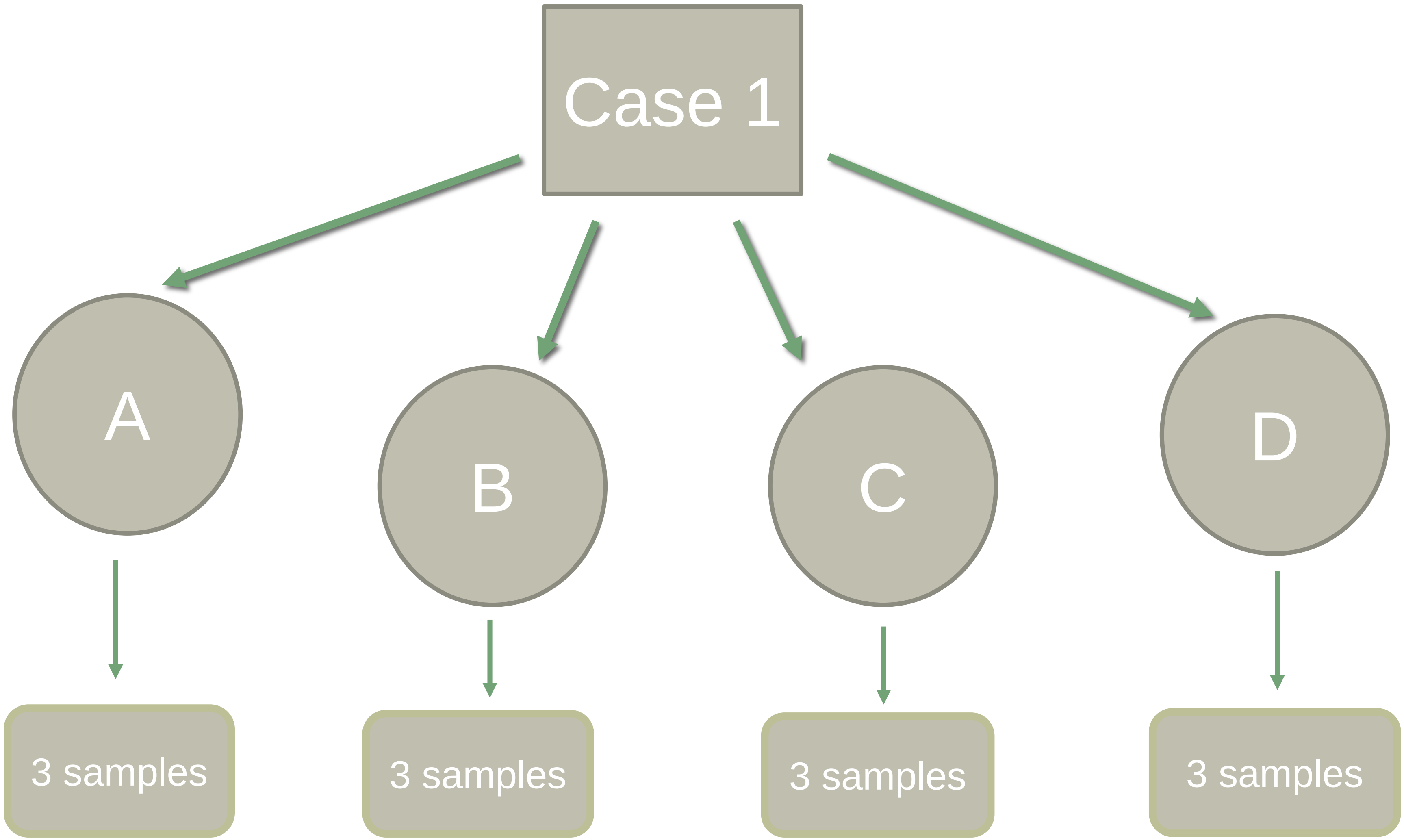
- Providing the forensic DNA community with a comprehensive analytical tool that will allow the genetic identification of *C. sativa* samples:
  1. Presence of clones
  2. Association between group of samples



MATERIALS AND METHODS

- Sampling (11 cases -199 samples)
- DNA Extraction
- DNA Quantitation
- 13 ISFG STR Multiplex
- Validation Studies
- Phylogenetic Analysis
- Population Database

Sampling (11 Cases)



Case 1

A

B

C

D

3 samples

3 samples

3 samples

3 samples

DNA Extraction

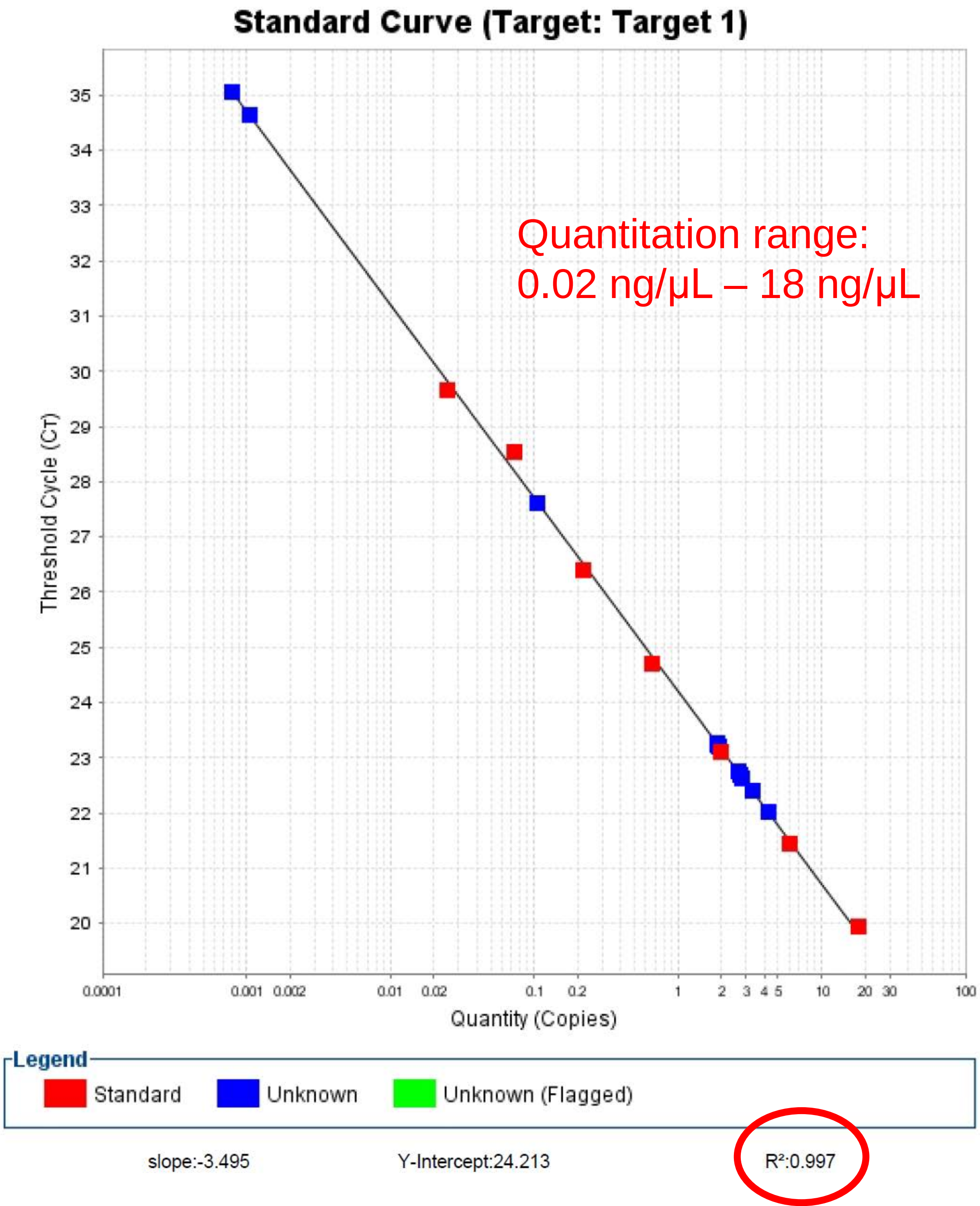
- DNeasy Plant Mini Kit (QIAGEN)
- Previously validated for marijuana DNA extraction (Coyle et al.)

DNA Quantitation

NEW!

- Development of real-time quantification system with marijuana specific primers (ANUCS304) (Howard *et al*.)
- Follow SWGDAM guidelines 9.4 and 9.5
- Results
  - Average: 3.47 ng/μL (Only 0.2 ng/μL needed for DNA profiling)
  - Standard deviation: ± 5.35 ng/μL

Quantitation



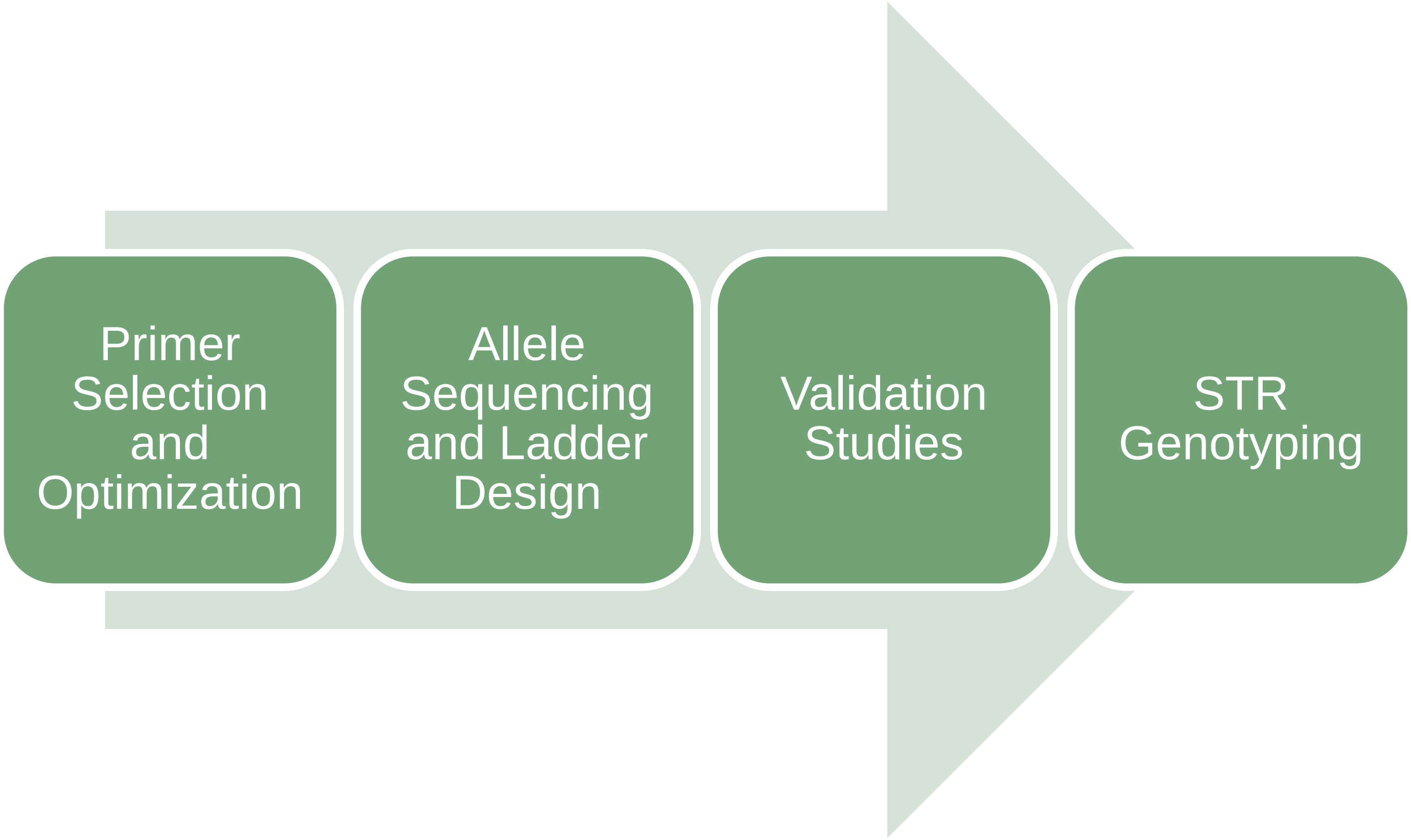
Quantitation range:
0.02 ng/μL – 18 ng/μL

slope:-3.495

Y-Intercept:24.213

R^2:0.997

13 ISFG STR Multiplex



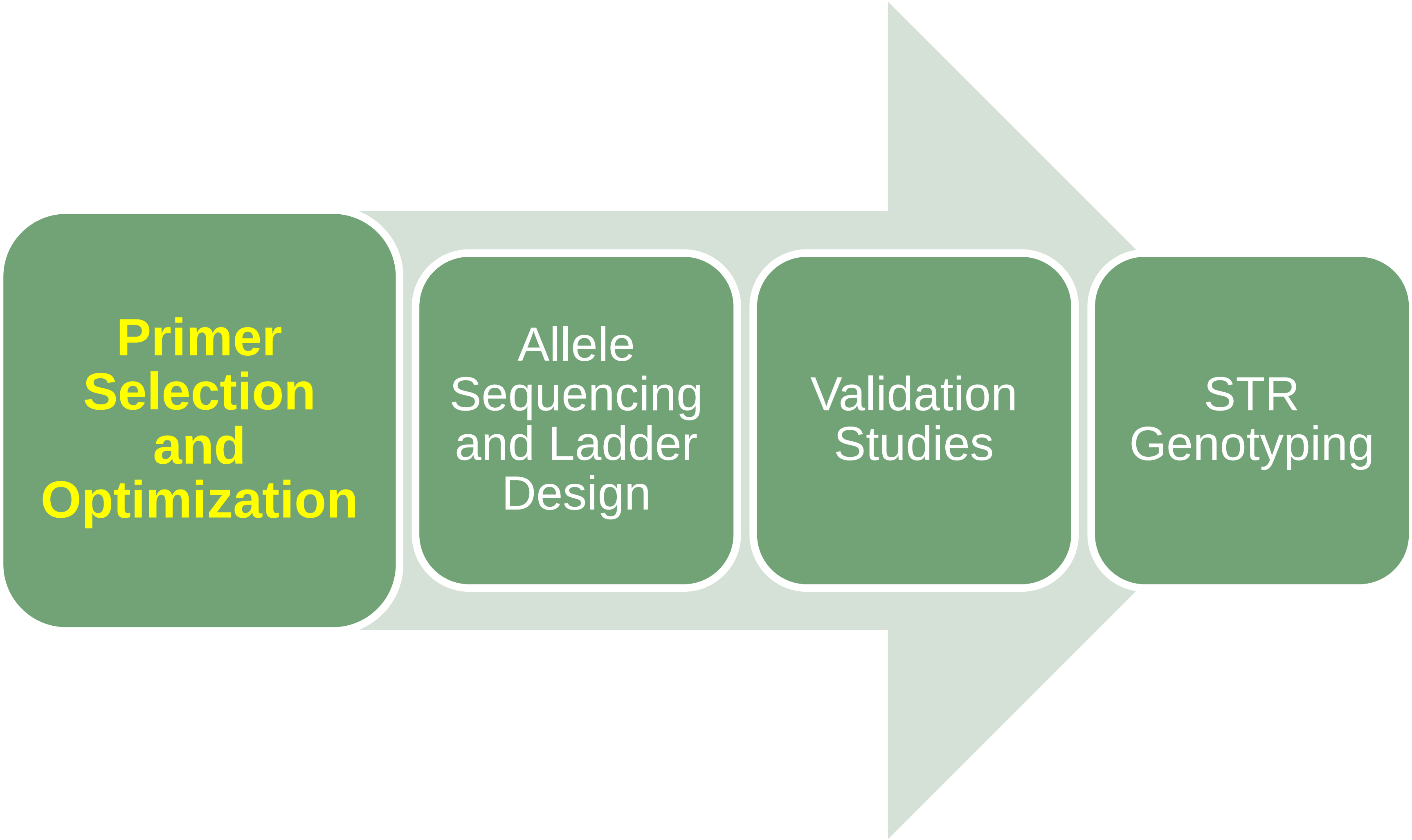
Primer Selection and Optimization

Allele Sequencing and Ladder Design

Validation Studies

STR Genotyping

13 ISFG STR Multiplex



Primer Selection and Optimization

Allele Sequencing and Ladder Design

Validation Studies

STR Genotyping

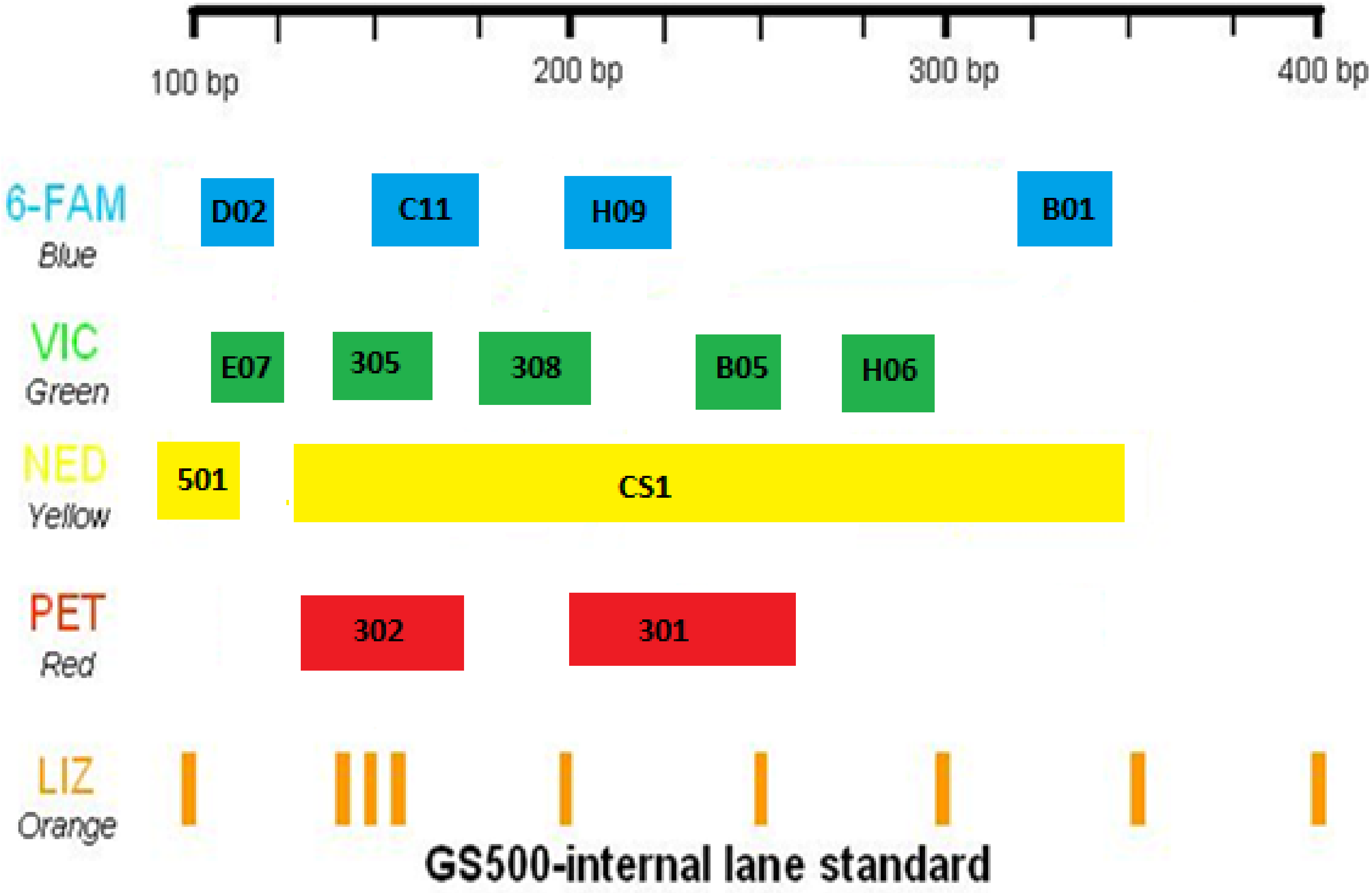
Primer Selection

Type	Name
Dinucleotide	H09, ANUCS308
Trinucleotide	E07, D02, B05, B01, ANUCS301, ANUCS305, H06, ANUCS302, C11
Pentanucleotide	ANUCS501
Hexanucleotide	CS1

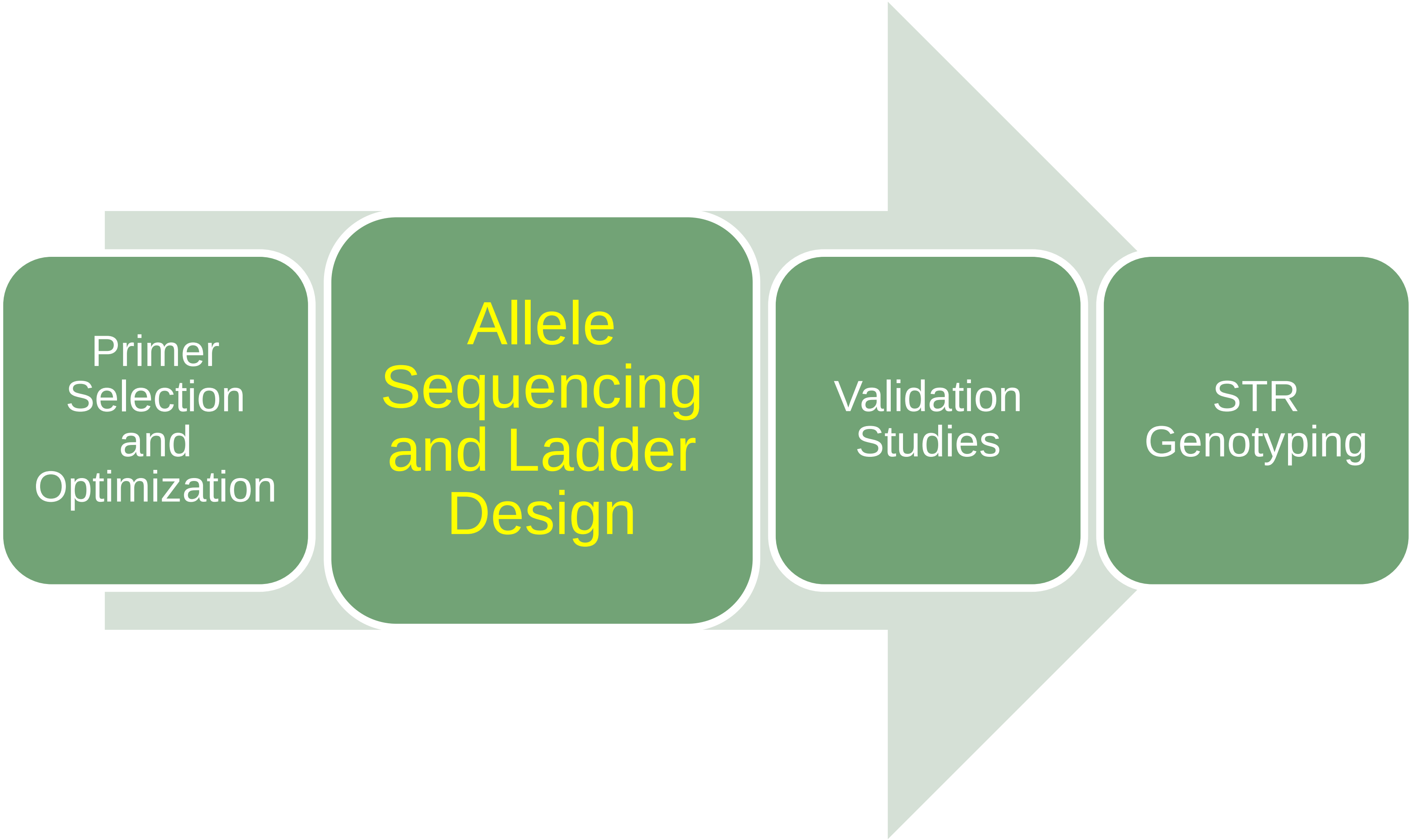
Multiplex Optimization

1. Multiplex manager software, using published forward and reverse primer sequences, to detect potential primer-primer interactions
2. Primer titration for multiplex
   - Type-IT Microsatellite Kit (Qiagen)

Final 13 Marker Multiplex



13 ISFG STR Multiplex



Primer Selection and Optimization

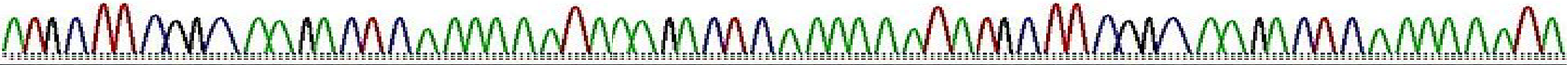
Allele Sequencing and Ladder Design

Validation Studies

STR Genotyping

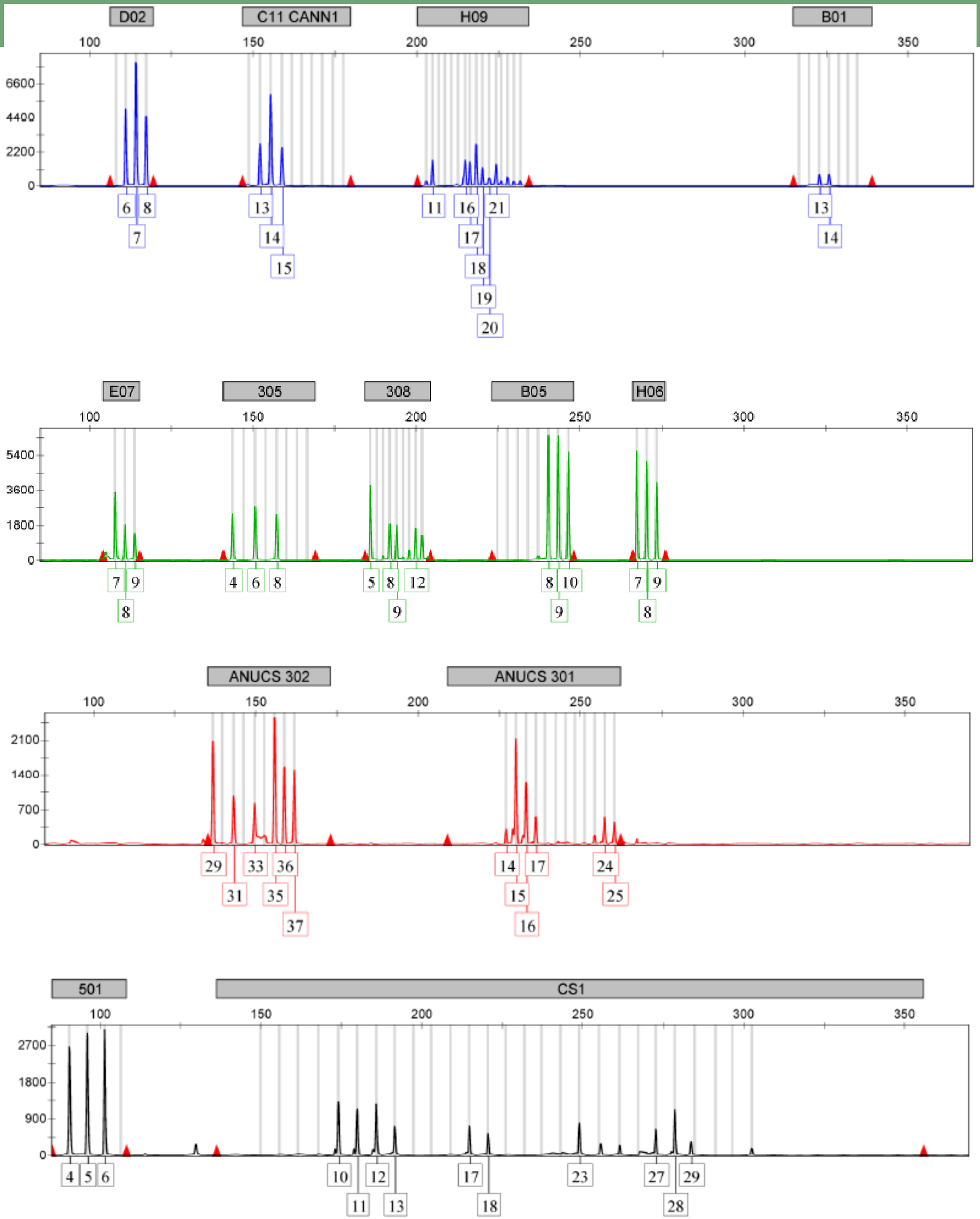
Sequencing and Allelic Ladder

- Two to eight alleles per marker
- Big Dye Direct Cycle Sequencing Kit
- CentriSEP purification columns
- Geneious Pro Software
- Allelic Ladder Design

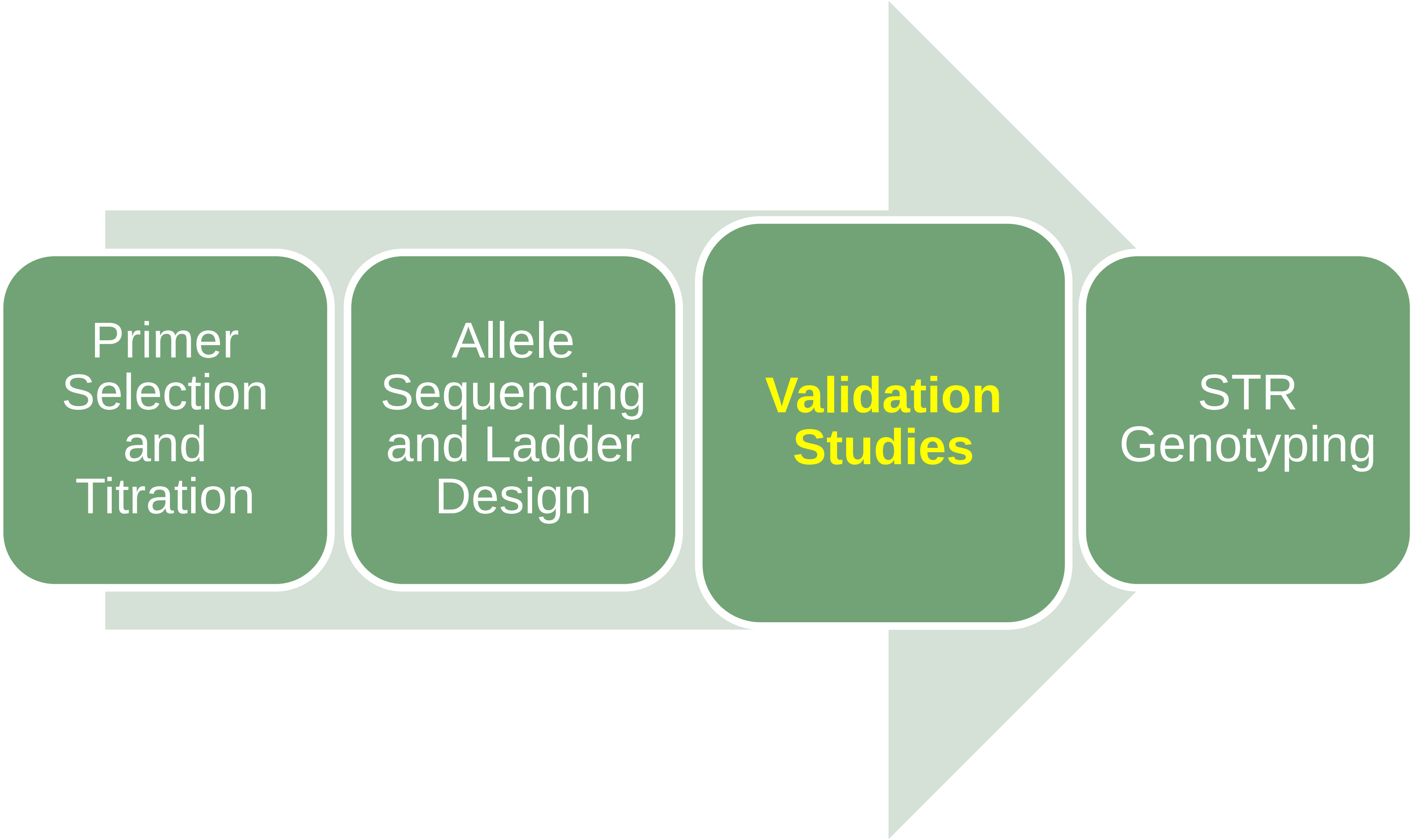


Marker	Fluorescent Label	Microsatellite Motif	Type of Repeat	Alleles
D02	FAM	(GTT)	Simple	6, 7, 8
C11	FAM	$(TGG)_x(TGA)_y(TGG)_z$	Compound/Indel	13, 14, 15, 21
H09	FAM	(GA)	Simple	11, 12, 13, 16, 17 18, 19, 20, 21, 23
B01	FAM	(AGA)	Simple	11, 13, 14, 15
E07	VIC	(ACT)	Simple	7, 8, 9
305	VIC	$(TGA)_x(TGG)_y(GGG)_z$	Compound	4, 5, 6, 8, 11
308	VIC	$(TT)_x(AT)_y$	Compound	5, 7, 8, 9, 12, 13
B05	VIC	(TTG)	Simple	3, 5, 7, 8, 9, 10
H06	VIC	$(AAC)_x(GAC)_y$	Compound	7, 8, 9
501	NED	(TTGTG)	Simple	4, 5, 6, 7
CS1	NED	(ATCACC)	Simple	6, 10, 11, 12, 13, 17, 18, 23, 24, 25, 26,27, 28, 29
302	PET	$(ACA)_x(ACA)_y(ACA)_z$	Compound/Indel	29, 31, 33, 35, 36, 37
301	PET	$(TAT)_x(GAT)_y$	Compound	15, 16, 17, 24, 25

Allelic Ladder



13 ISFG STR Multiplex



Primer Selection and Titration

Allele Sequencing and Ladder Design

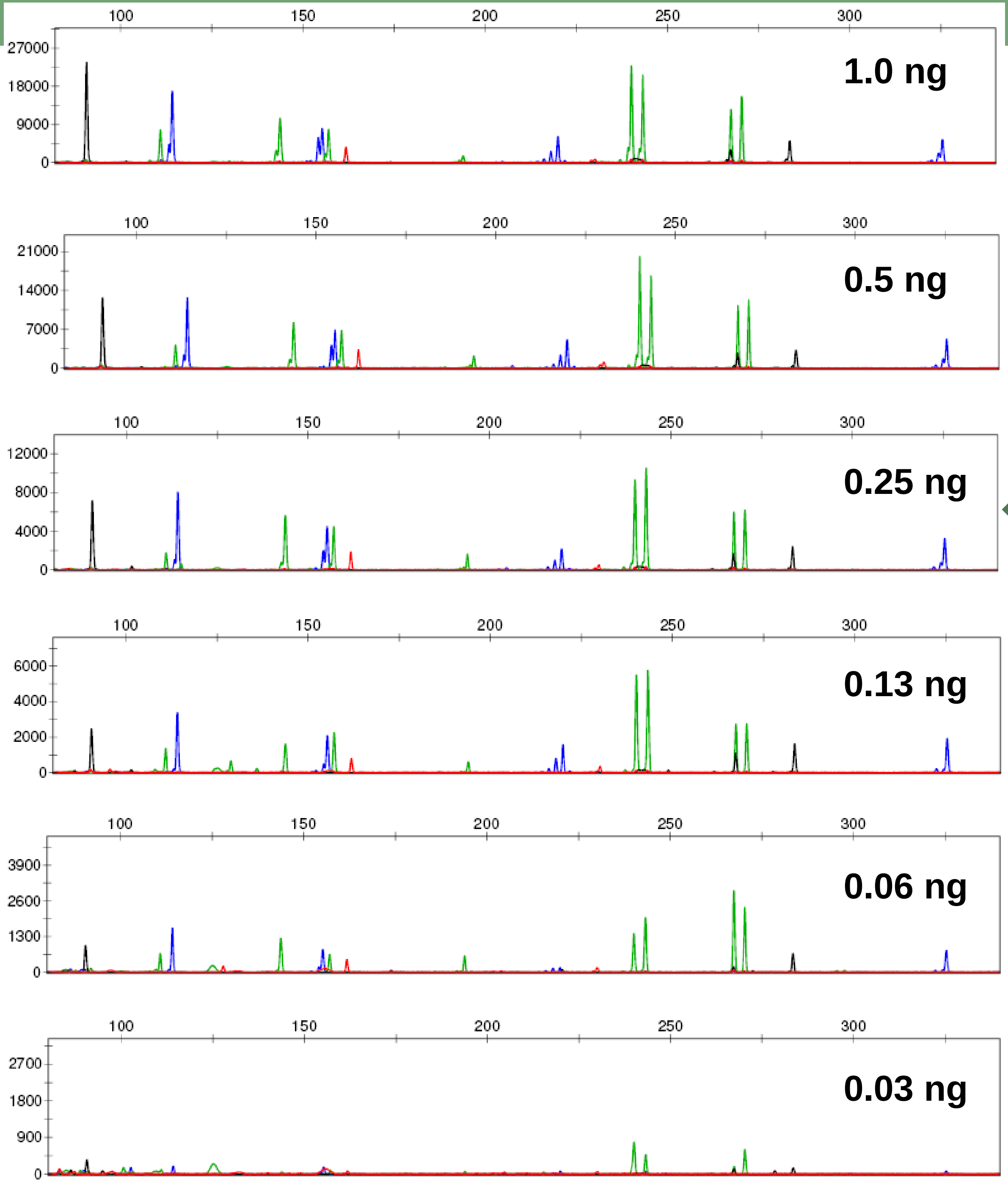
Validation Studies

STR Genotyping

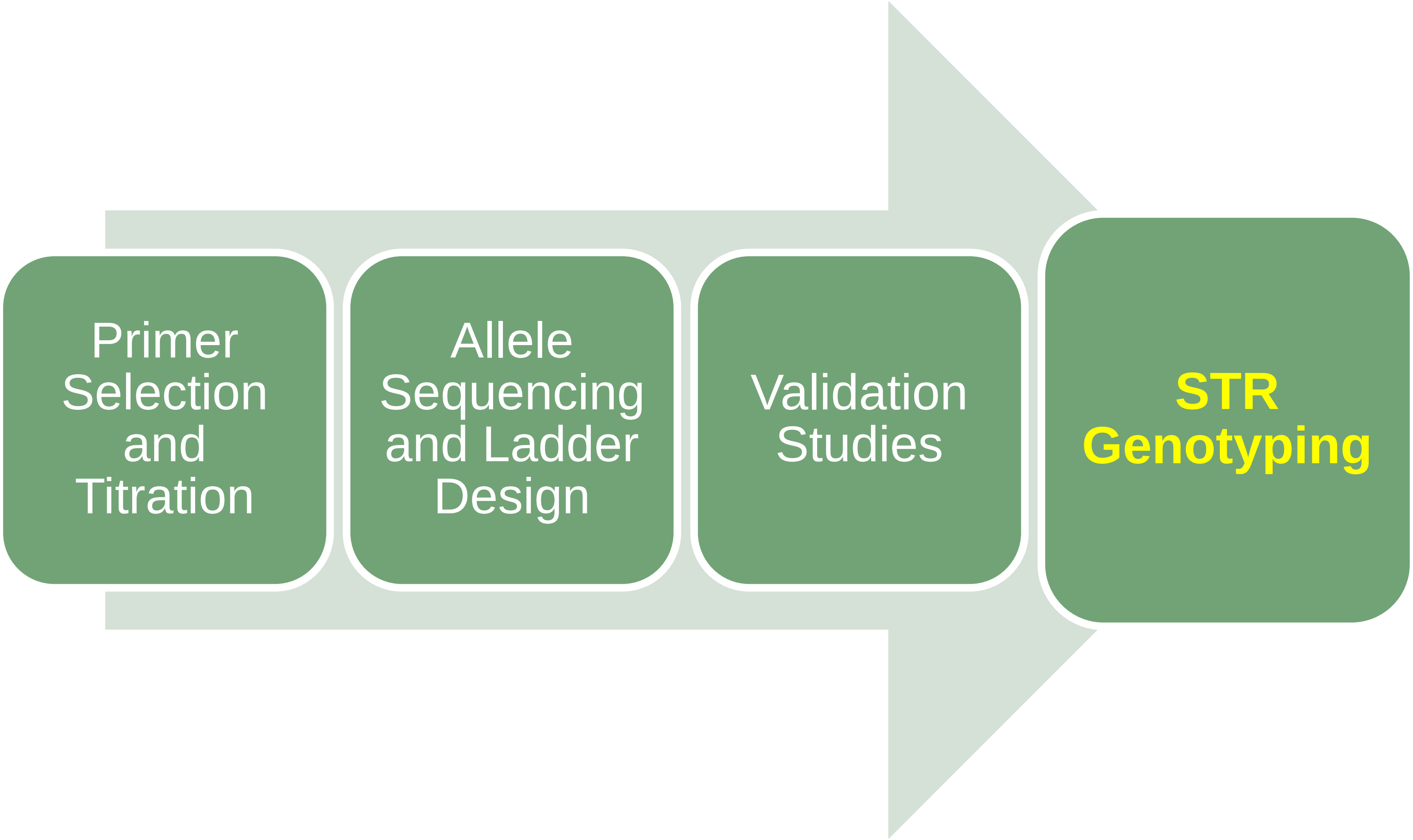
Validation Studies

- The optimized developed system was validated according to SWGDAM guidelines:
  - Sensitivity
    - Dynamic range of assay (1.0 ng – 31.2 pg)
  - Species specificity
    - Cross-reactivity
      - Plants (*Humulus lupulus* and *Nicotiana tabacum*)
      - Animals
      - Human

Sensitivity



13 ISFG STR Multiplex



Primer Selection and Titration

Allele Sequencing and Ladder Design

Validation Studies

STR Genotyping

STR Genotyping

Previously Extracted Samples (N =199) used the
Dneasy Plant Mini Kit

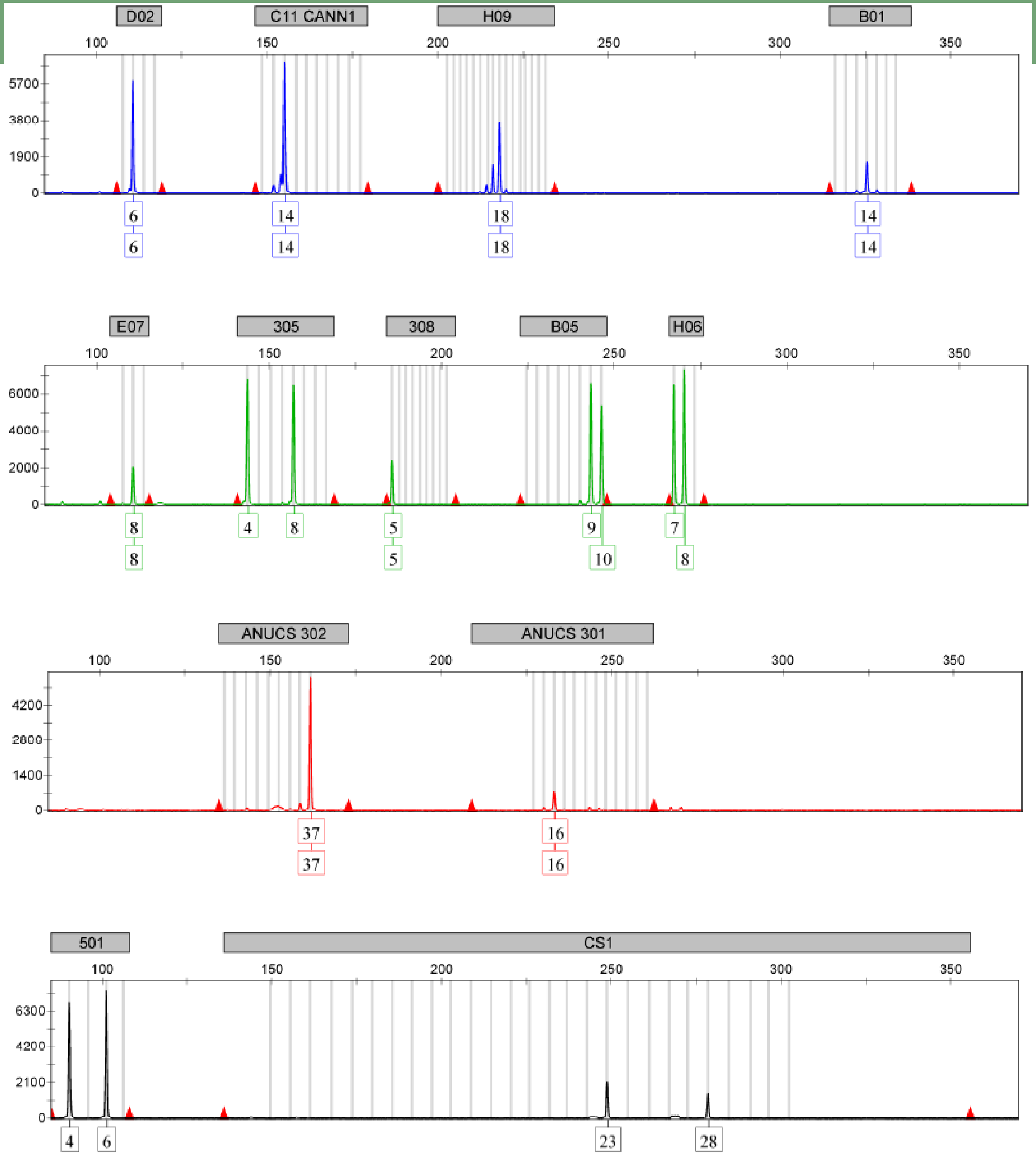
DNA Quantitation
SYBR Green on StepOne Real Time PCR Instrument

PCR
Type-It Microsatellite Kit on Ependorf MasterCycler Gradient

Electrophoresis
on ABI 3500 Genetic Analyzer

Genemapper v 4.1 analysis

Positive Control

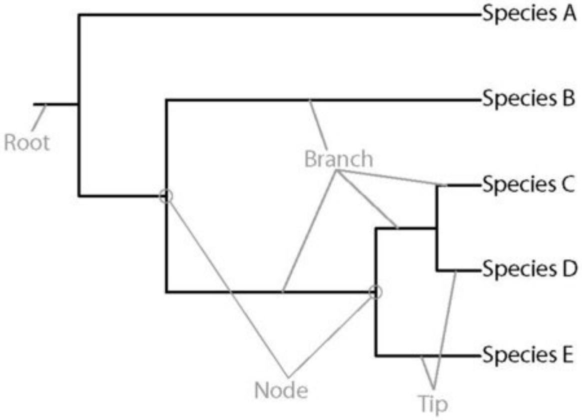


STR Results

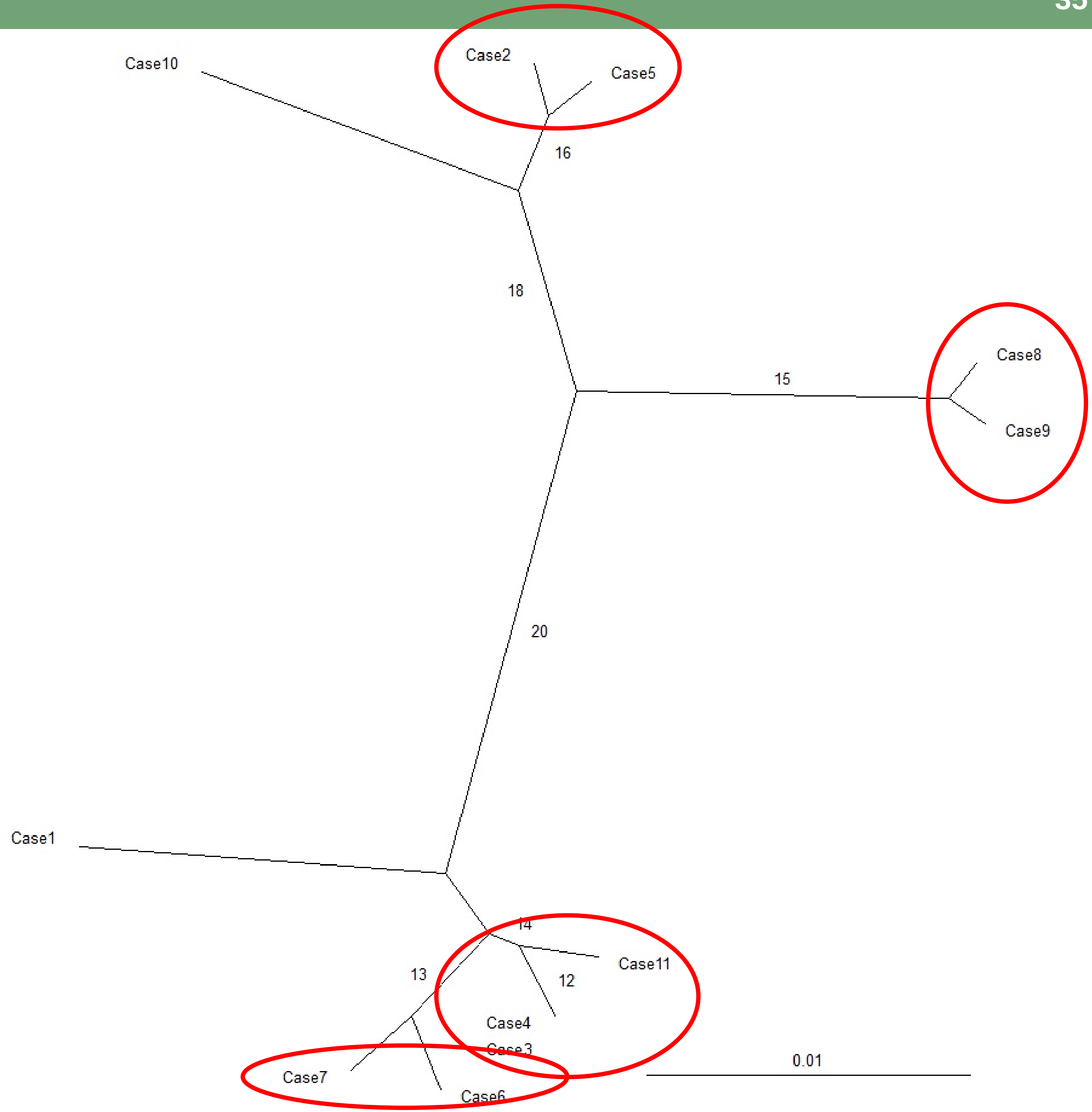
- Determine number of multi-locus genotypes and genotype sharing among samples
- Results:
  - Only 4 duplicate genotypes within seizures were found
- 127 samples generated full profiles (64%)
  - 36% partial profiles
    - Locus dropout: ANUCS308, ANUCS302, ANUCS301, B01

Phylogenetic Analysis

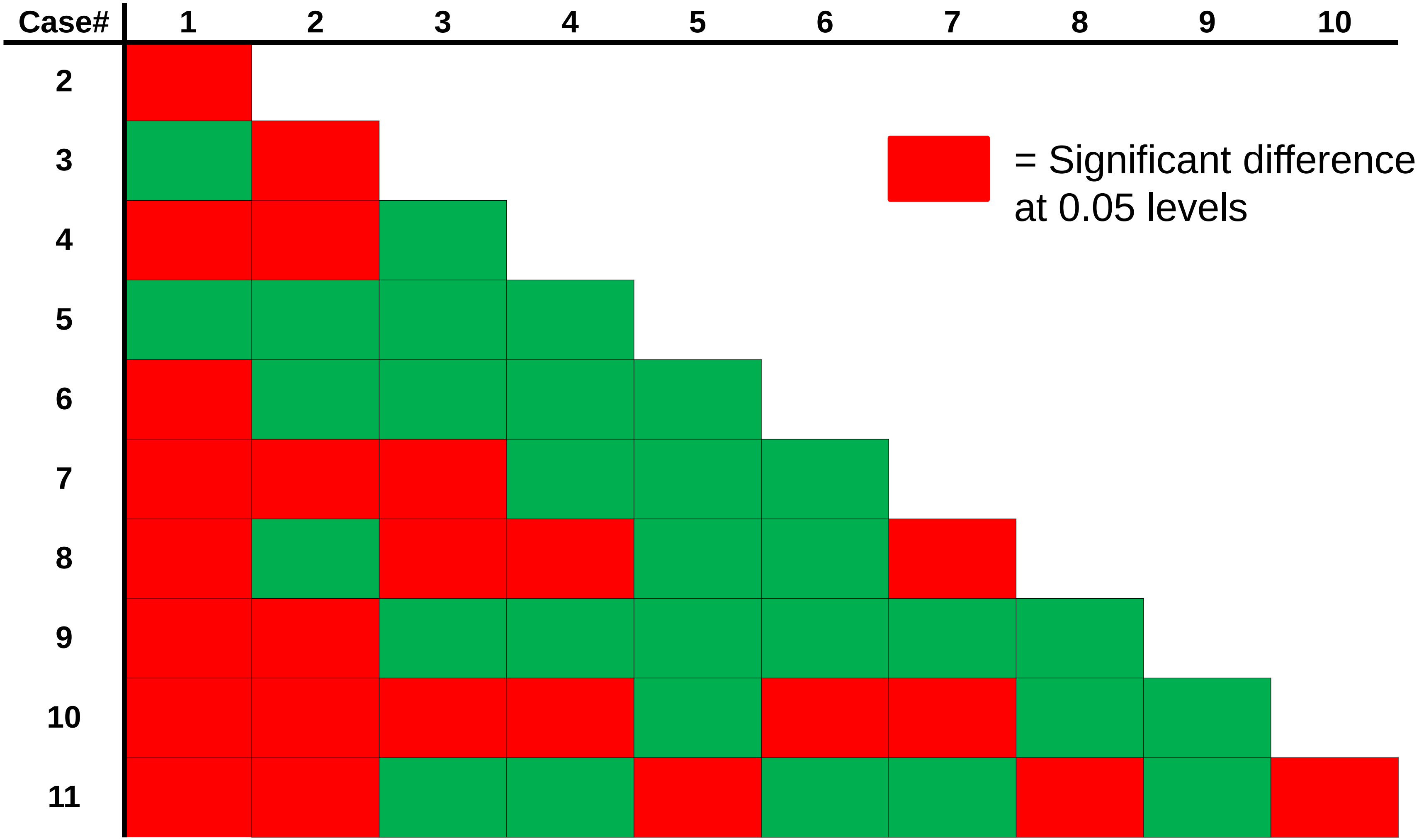
- UPGMA method with coefficient of ancestry *Fst* as genetic distance;
    - GDA software
- Genetic association confirmed with Arlequin software
- Screen for the presence of homogeneous sub-populations among samples



UPGMA Method, GDA



Arlequin Software, *Fst*

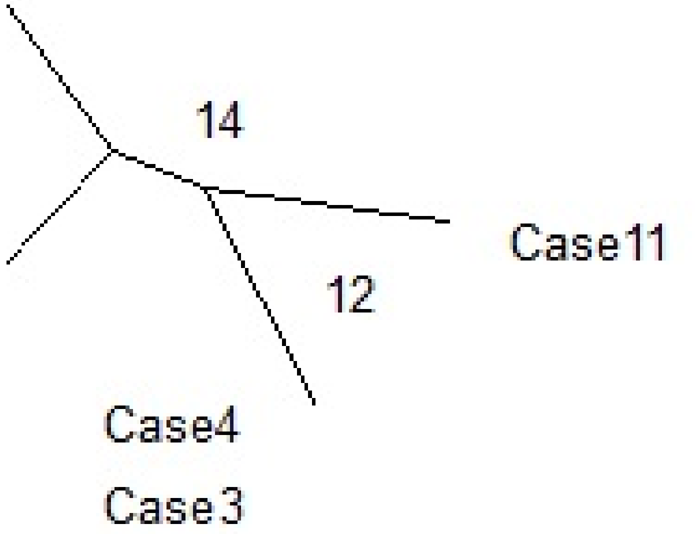


Case#	1	2	3	4	5	6	7	8	9	10
2	🟥									
3	🟩	🟥								
4	🟥	🟥	🟩							
5	🟩	🟩	🟩	🟩						
6	🟥	🟩	🟩	🟩	🟩					
7	🟥	🟥	🟥	🟩	🟩	🟩				
8	🟥	🟩	🟥	🟥	🟩	🟩	🟥			
9	🟥	🟥	🟩	🟩	🟩	🟩	🟩	🟩		
10	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟩	🟩	
11	🟥	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟥

🟥 = Significant difference at 0.05 levels

Homogenous sub-population

- Genetic association of three pairs of cases.
  - Subset of samples (N=97) was found to form a homogeneous subpopulation (low *Fst*) in Hardy – Weinberg equilibrium and linkage equilibrium (GDA).



GDA: 95% confidence interval bootstrapping
Fst = 0

Reference Population Database

- Population genetic statistics (PowerStats v1.2) and parameters of forensic interest with homogeneous sub-populations
  - Allele Frequencies
  - Hardy-Weinberg Equilibrium
    - Observed heterozygosity
    - Expected heterozygosity
  - Screen for null alleles using the GenePop software
  - Random match probability

Parameters of Forensic Interest

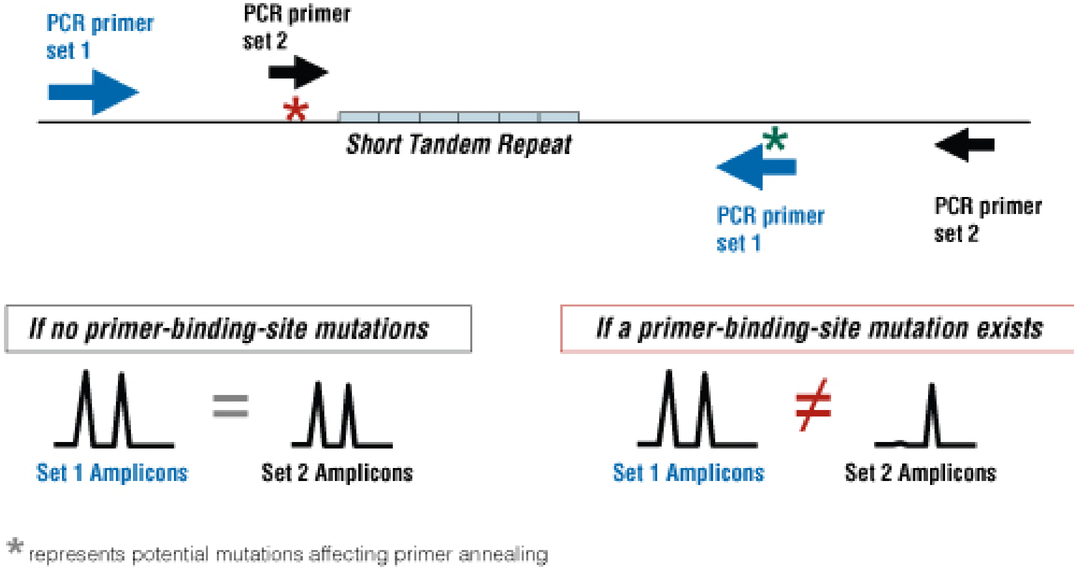
	D02-CANN1	C11-CANN1	H09-CANN2	B01-CANN1	E07-CANN1	ANUCS 305	ANUCS 308	B05-CANN1	H06-CANN2	ANUCS 501	CS1	ANUCS 302	ANUCS 301
Ho	0.567	0.417	0.630	0.226	0.547	0.563	0.155	0.516	0.515	0.438	0.753	0.389	0.269
He	0.566	0.441	0.715	0.522	0.656	0.590	0.312	0.551	0.520	0.513	0.818	0.565	0.609
PIC	0.47	0.36	0.69	0.41	0.58	0.5	0.28	0.47	0.43	0.4	0.78	0.52	0.54
PD	0.718	0.61	0.887	0.664	0.82	0.748	0.431	0.733	0.676	0.671	0.932	0.738	0.746

Ho: Observed heterozygosity
He: Expected heterozygosity
PIC: Polymorphic information content
PD: Power of Discrimination

The combined power of discrimination of this multi-locus system was 1 out of 70 million.

Null Allele Analysis, Genepop

- Null alleles detected in ANUCS308, ANUCS302, ANUCS301, B01
- Primer-binding interactions
  - Need to design primers
- Corrected allele frequency



Conclusions

1. Development of a real-time PCR quantitation system (calibration curve showing $r^2 = 0.99$ + melting curve showing primer specificity)
2. Evaluation of 13-loci STR multiplex + ISFG recommendations
3. Detection of null alleles
4. Phylogenetic tree
5. Reference population database

POTENTIAL IMPACT

Future Studies

Cannabis sativa samples (expansion from 200 to 400 individuals)

13 ISFG STR Multiplex
With new tetranucleotides

Organelle DNA Typing

STR DNA Typing
MPS Analysis

Acknowledgments

- Special thanks to:
  - Sam Houston State University
  - Dr. Sheree Hughes-Stamm
  - Dr. David Gangitano
  - Lori Powe

Questions?