

Development of a Next Generation Sequencing Panel Targeting Cannabinoid Synthase Genes to Distinguish Between Marijuana and Hemp

Ya-Chih (Jessica) Cheng, MS., Rachel Houston, PhD.

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



Role of Funding

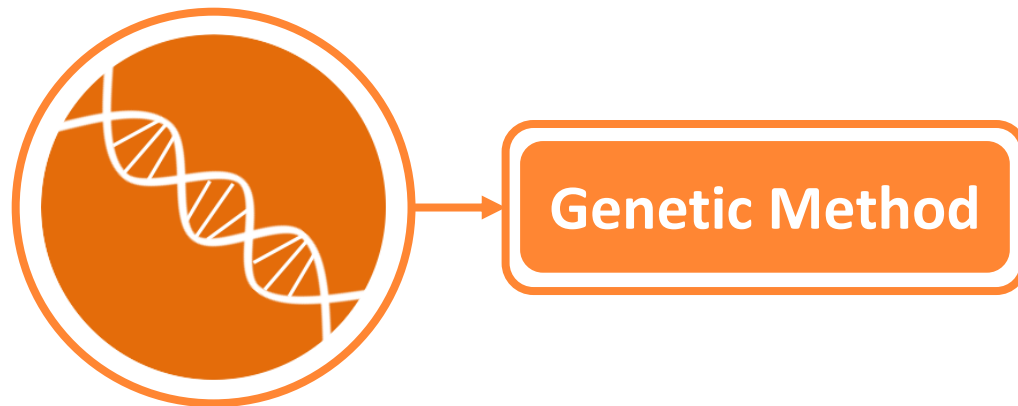
- This project was funded by an award to Rachel Houston, PhD at Sam Houston State University through the Center for Advanced Research in Forensic Science (CARFS), an NSF-funded IUCRC at Florida International University (Award # 1739805).
- The opinions, finding, conclusions, or recommendations expressed in this presentation are those of the authors.

Differentiation of Marijuana and Hemp-Statement of Problems



➤ Traditional Method Limitations

- Microscopic, Color test → cannot differentiate hemp and marijuana
- Chemistry instrumentation → provide limited information, effected by age, storage, sample type and size



➤ Genetic Method

- Only need trace amount of sample available
- Analyzed non-traditional sample types
- Not effect by sample condition or storage
- Crop type differentiation
- Tracing geographic origin

Genetic Markers for Differentiating Hemp and Marijuana

Int J Legal Med (2016) 130:635–647



Article

Developing and (Hemp) for The

TECHNICAL NOTE GENERAL

Cristiane Barbosa D'Oliveira Matos
M.S.; Lilian de Oliveira Machado,
Dobler,⁴ M.S.; Renata Machado
Roesch,⁴ Ph.D.; Flávio Anastácio

Whole Plastome Sequencing of
Drug-Type *Cannabis*: Insights into
Use of Plastid in Forensic Analyses



Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Forensic Science International 159 (2006) 132–140

www.elsevier.com/locate/forsciint

Forensic
Science
International

DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in “drug-type” and “fiber-type” *Cannabis sativa* L.

Mareshige Kojoma^{a,b,*}, Hikaru Seki^b, Shigeo Yoshida^b,
Toshiya Muranaka^b

^a JYUGEI Institute (The Arboricultural Research Institute), University Forests,
Graduate School of Agricultural and Life Sciences, The University of Tokyo, 457 Kano,
Minami-izu, Shizuoka 451-0304, Japan

^b Plant Science Center, RIKEN, 1-7-22 Suehirocho, Tsurumi-ku, Yokohama 230-0045, Japan

Received 20 May 2005; received in revised form 9 July 2005; accepted 18 July 2005

Available online 6 September 2005

Madeline G. Roman • David Gangitano • Rachel Houston

Received: 25 June 2019 / Accepted: 6 August 2019

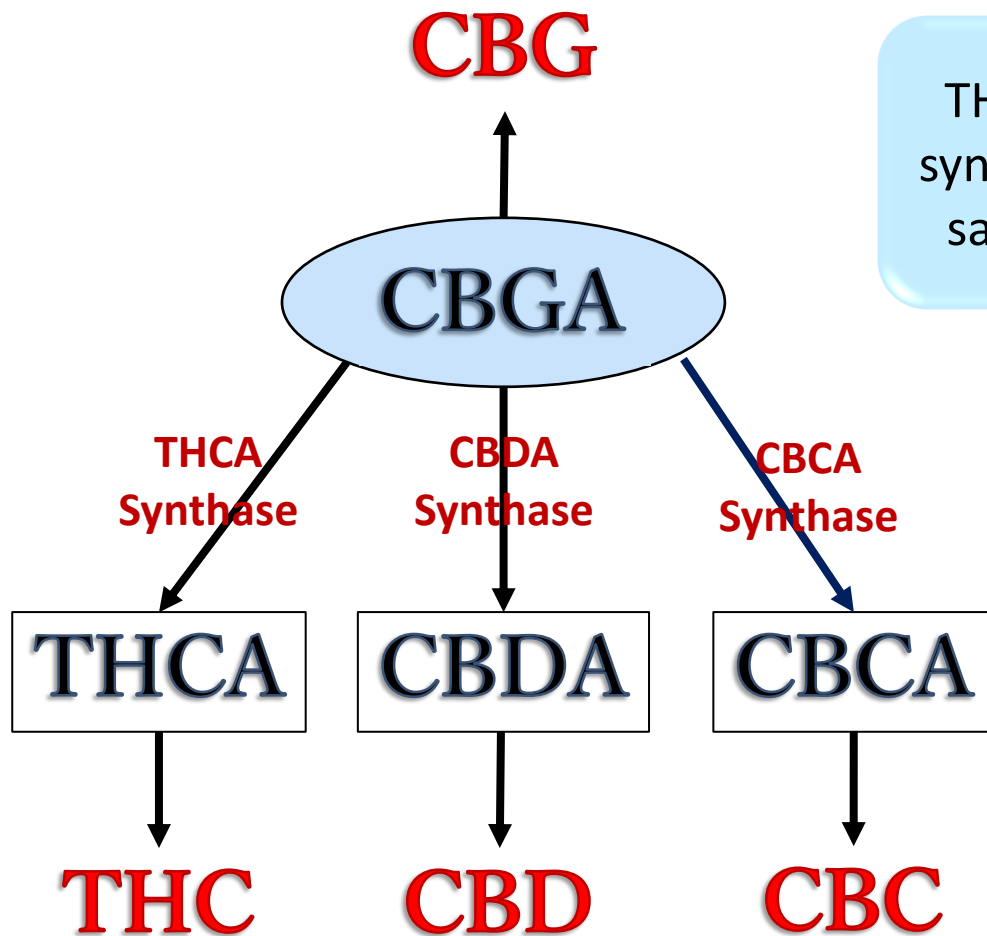
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Legal Medicine 26 (2017) 33–40



Synthase Gene and Chemotype

Cannabis strains are usually differentiated by their chemical composition (chemotype)



THCA, CBDA, and CBCA synthase compete for the same substrate (CBGA)

The genotypes differences of the three synthase genes might influence cannabinoid content
→ Differentiation Markers

➤ Synthase genes sequences

- CBCAS is reported to be 96% identical to THCAS sequence
- Cannabinoid synthase genes copies may include functional genes as well as several pseudogenes
- Pseudogene with 89%– 95% similar in identity to THCAS, CBDAS, and CBCAS

Study's Goal and Approach



➤ Limitation of Current Genetic method

- Need to incorporated CBCAS and pseudogene
- Limited sample varieties, small sample number (FP and FN still occurred)
- Lack of a comprehensive *C. sativa* database

➤ Goal

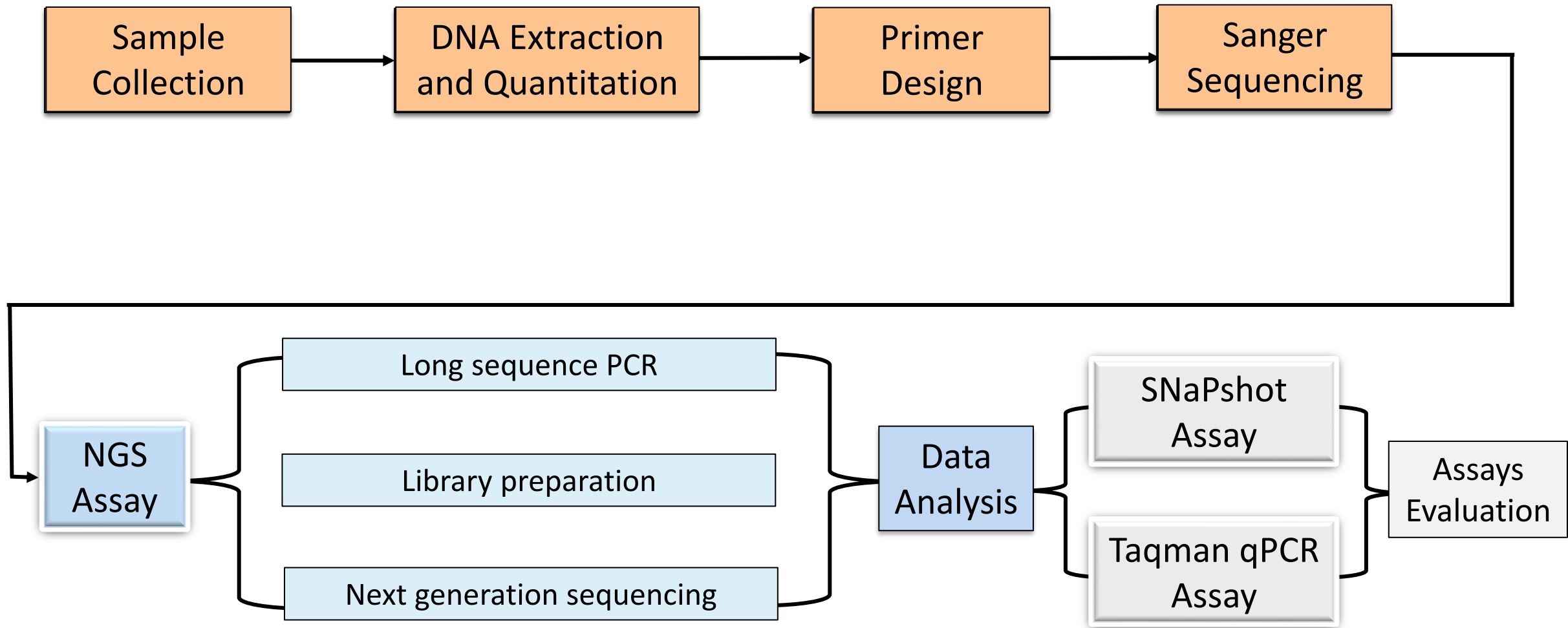
Provide large scale evaluation of the synthase genes and pseudogenes on wide varieties of Cannabis to identify novel polymorphisms or pattern of genes' present for crop type discrimination.



➤ Designed NGS panel targeting three synthase genes and pseudogenes

- NGS can target specific genes of interest with high coverage and confidence
- Quickly screen and sequence several types of cannabis with its high-throughput capabilities

Flow Chart



Cannabis Sample Collection

	Crop Types	Number	Source	Chemistry Data
1	Reference Marijuana	27	NIDA, NIST	Quant & TIC Data
2	Reference Hemp	3	NIDA, NIST, University of Kentucky	Quant & TIC Data
3	Seized Marijuana	61	DEA (University of Mississippi)	Quant Data
4	CBD Hemp	20	Online purchase	Quant & TIC Data
5	CBG Hemp IH	13	University of Mississippi	Not Available
	CBG Hemp	8	Online purchase	Quant & TIC Data
6	Δ 8-THC Hemp	7	Online purchase	Quant & TIC Data
7	THCO Hemp	3	Online purchase	Quant & TIC Data
8	Δ 10-THC Hemp	1	Online purchase	Quant & TIC Data
9	CBDV Hemp	1	Online purchase	Quant & TIC Data
10	Hemp Seed	4	Grocery Store purchase	Not Available
	Total	148	Marijuana: 88	Hemp: 60

Primer Design

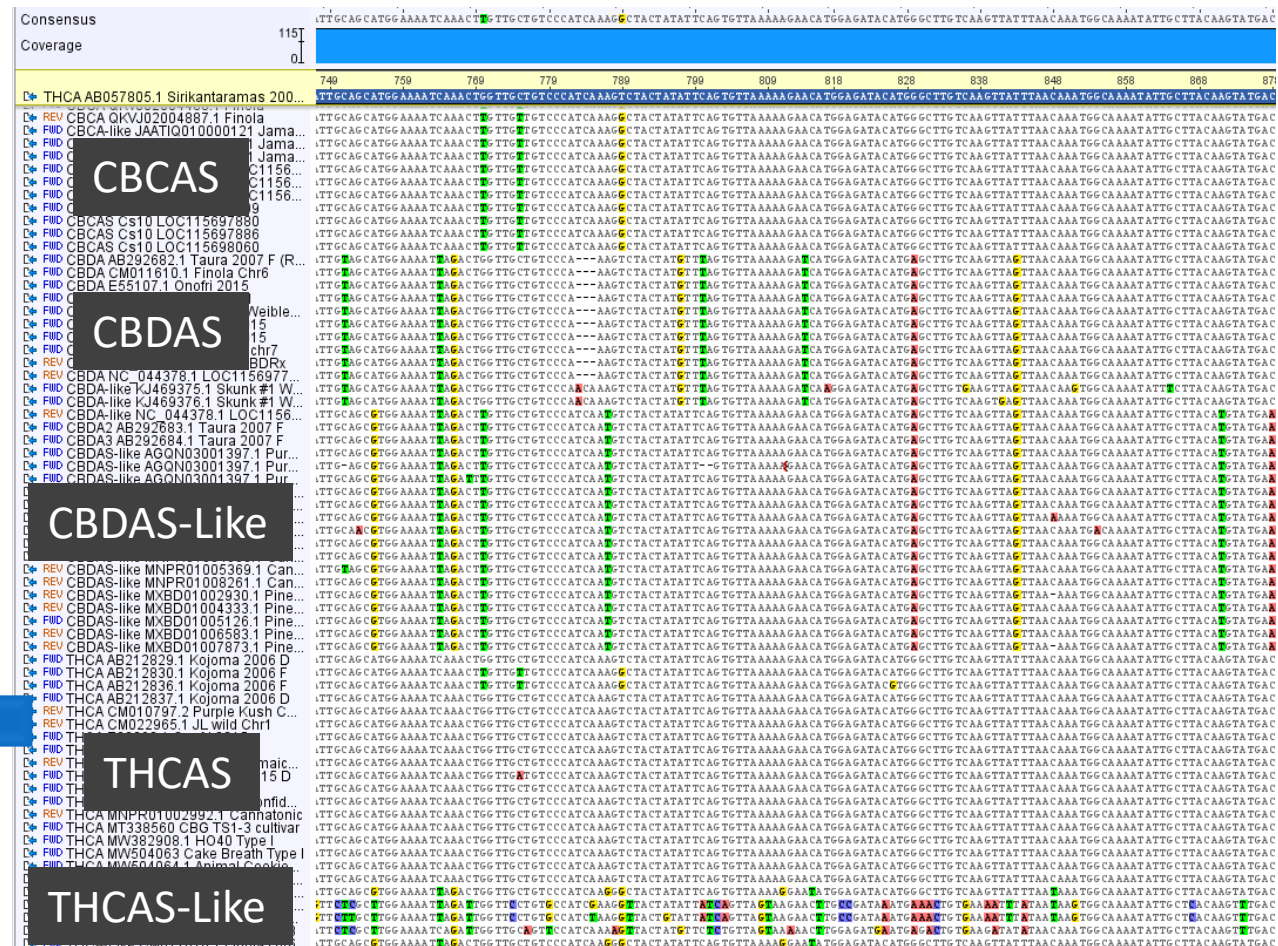
Approach 1: Using Published Primers

Name	Primer	Reference
THCAS F2	ACTGAAGAAAAATGAATTGCTCAG	Cascini 2019
THCAS R2	ATTTAAAGATAATTAATGATGATGCG	Cascini 2019
CBDAS F1	ATGAAGTGCTCAACATTCTC	Cascini 2019
CBDAS Lav F	CTGCAGGAATGAAGTACTCAACATTCTCCTTTTGG	Laverty 2019
CBDAS R1	TTAATGACGATGCCGTGG	Cascini 2019
CBDAS Lav R	AAGCTTTTCATGGTACCCCATGATGATGCCGTGGAAGAG	Laverty 2019
CBCAS Lav F	CGGATGTACTGTTATGCTCCAA	Laverty 2019
CBCAS Lav R	CATTCTCCATTAAATAAGAAAGACAA	Laverty 2019

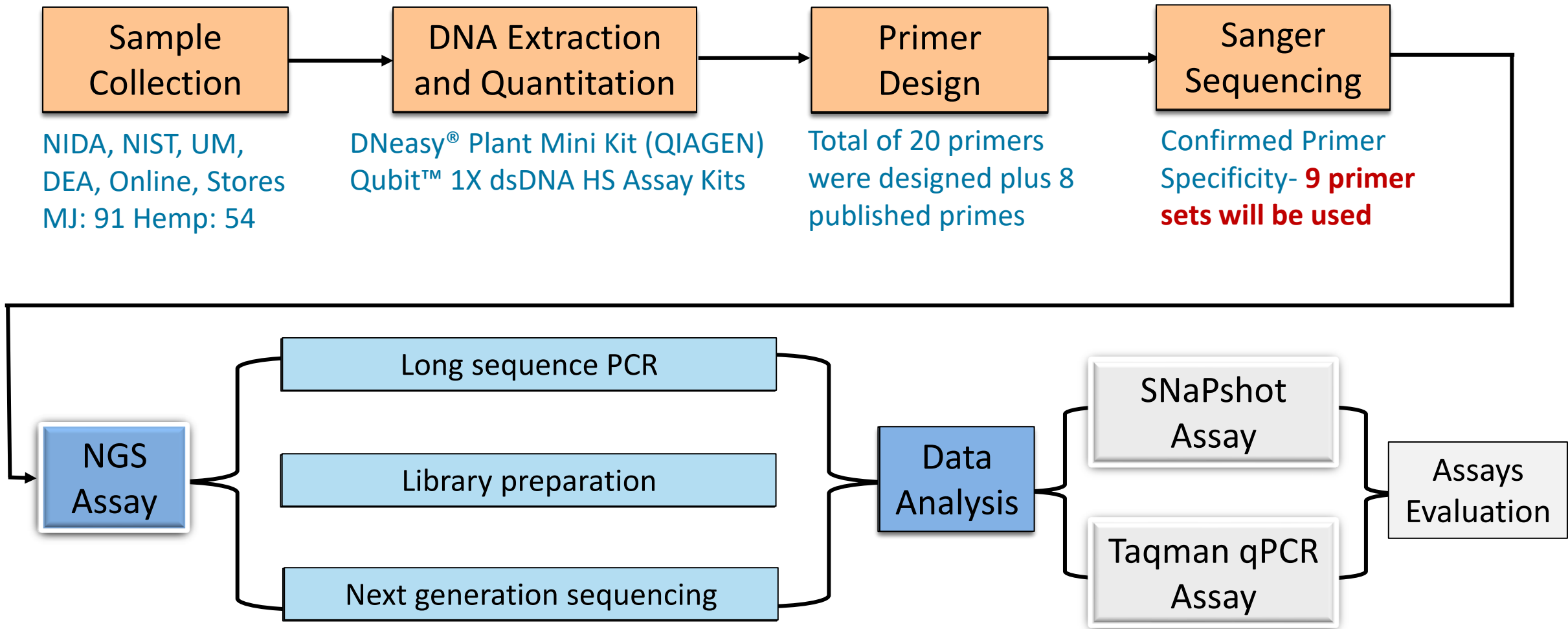
No.	Primer Name	No.	Primer Name	No.	Primer Name
1	THCAS F4	8	CBCAS R2	15	CBDA+Like F2
2	THCAS R4	9	THCA+Like R	16	CBDA+Like F3
3	CBDAS F3	10	Everyone R1	17	CBCA+Like F
4	CBDAS R3	11	CBDA+Like R1	18	THCA F5
5	CBCAS F1	12	THCA R5	19	Everyone R2
6	CBCAS F3	13	CBDA+Like R2	20	Everyone F1
7	CBCAS R1	14	CBDA+Like F1		

Approach 2: Aligned Published Syntheses' Sequence

Align 121 Reference Sequences (including THCAS, THCAS-like, CBDAS, CBDAS-like, CBCAS)



Flow Chart



General Workflow

Long-Range PCR

- LA PCR Kit Ver.2.1
- Single-plex PCR

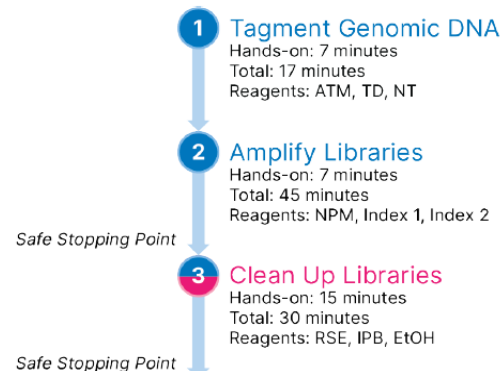


PCR Product Pool

- Qubit quantitation
- PCR Products Pool (0.2 ng/μL)
- Pool all targets from the sample
- ExoI clean-up

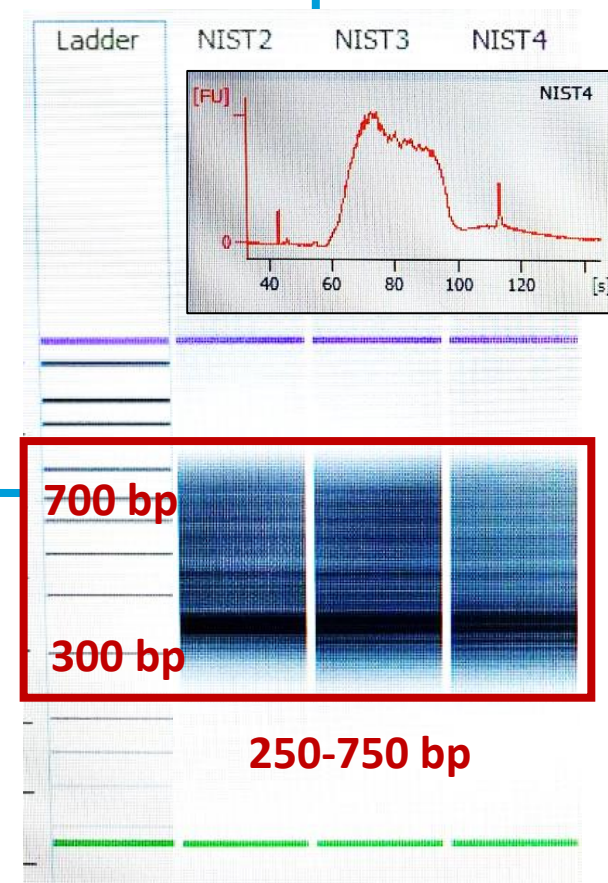
Library Preparation

Figure 1 Nextera XT DNA Library Prep Workflow



Libraries Quality Check

- Bioanalyzer



MiSeq FGx Run

- Denature libraries
- Dilute denature libraries
- Instrument loading

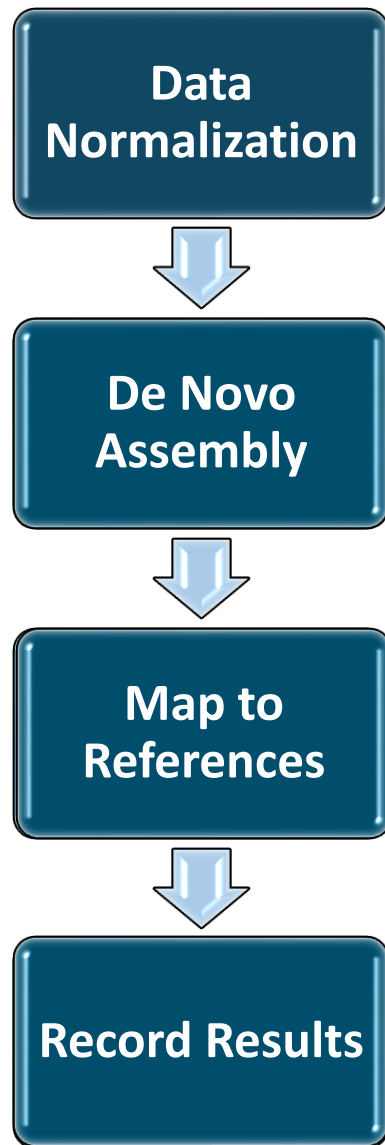


Libraries Normalization

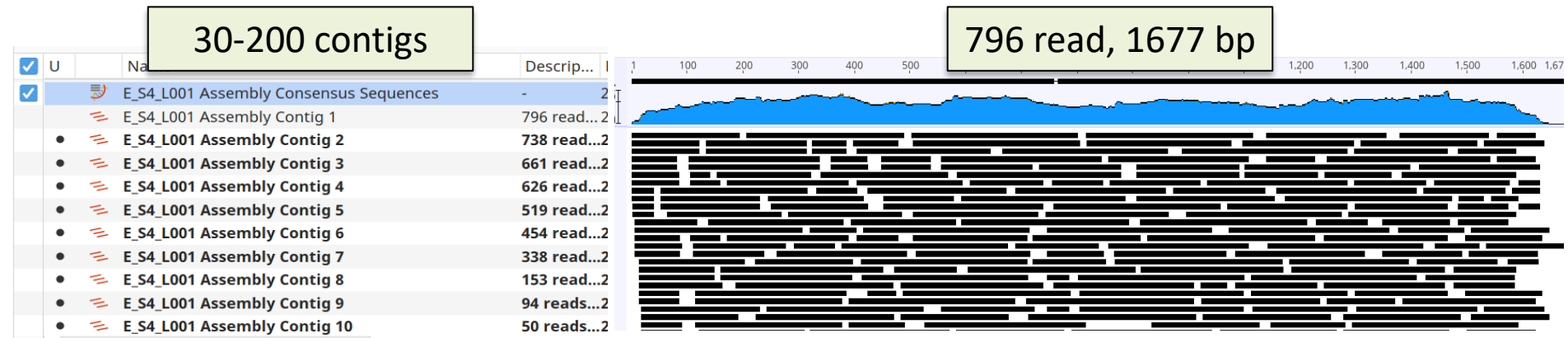
- Qubit quantitation
- Libraries Pool (1.32 ng/μL)

$$\text{Molarity} = \frac{\left(\frac{\text{ng}}{\mu\text{l}}\right)}{660 \frac{\text{g}}{\text{mol}} \times 500\text{bp}}$$

Data Analysis-Geneious Prime



Assemble millions of reads generated by NGS platforms: align and merge overlapping fragments of a DNA sequence to reconstruct the original sequence.



Map the contig to Reference Sequence: assemble contig sequences to THCAS, CBDAS, and CBCAS sequences to separate pseudo genes and locate differences

Excel Sheet: record present or absent of the synthase genes and pseudo synthase genes, as well as the polymorphisms on the synthase genes

228 documents assembled to 3 reference

- ✓ 122 of 228 reads were assembled to 3 reference sequences to produce 3 contigs
- ✗ 106 reads were not assembled

Assembled 231 documents using Geneious assembler. [Show Options](#)
Assembly Duration: 8.71 seconds (23.28 seconds CPU time)
Performed 5 of up to 5 iterations.
First iteration matched 111 reads.
Best iteration matched 122 reads.

Sample	Strain	THC	CBD	CBC	THC/CBD	THCA	THCA:Line 1	THCA:Line 2	THCA:Line 3	THCA:Line 4	CBD:Line 1	CBD:Line 2	CBD:Line 3	CBD:Line 4
DEABIA	Strain M1 2	1.514	1.121	0.089	1.36	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
DEABIA	Strain M1 2	9.967	0.074	0.195	134.65	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
DEABIA	Strain M1 2	2.049	0.009	0.185	227.67	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
DEABIA	Strain M1 2	0.199	0.006	0.068	56.50	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
DEABIA	Strain M1 2	2.407	0.01	0.148	260.70	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
DEABIA	Strain M1 2	0.125	0.003	0.098	108.33	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
DEABIA	Strain M1 2	0.074	0.004	0.055	0.06	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
DEABIA	Strain M1 2	0.71	0.004	0.067	182.50	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H10	Strain 2	0.64	12.96	0.02	0.049	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H11	Strain 2	0.64	12.96	0.02	0.049	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H12	Strain 2	0.79	16.41	0.02	0.049	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H13	Strain 2	0.76	13.63	0.04	0.056	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H14	Strain 2	0.24	20.137	0.17	0.012	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H15	Strain 2	0.51	19.489	0.163	0.026	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H16	Strain 2	0.24	21.814	0.188	0.013	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H17	Strain 2	0.52	18.269	0.144	0.028	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H18	Strain 2	0.24	18.802	0.17	0.016	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H19	Strain 2	0.22	17.121	0.258	0.013	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H20	Strain 2	0.51	18.284	0.104	0.028	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H21	Strain 2	N/A	N/A	N/A	N/A	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H22	Strain 2	N/A	N/A	N/A	N/A	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
H23	Strain 2	0.24	17.264	0.178	0.016	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

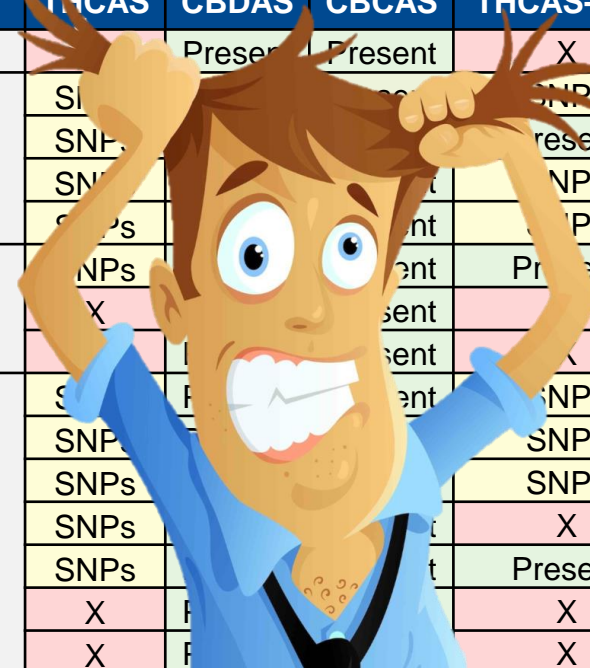
NGS Results Summary (Partial)

Sample	Strain	THCAS	CBDAS	CBCAS	THCAS-Like	CBDAS-Like
H18	CBD Hemp	X	Present	Present	X or SNPs	X
H20		X	SNPs	Present	X or SNPs	X
H24		X	Present	Present	X or SNPs	X
H14		X	Present	Present	X or SNPs	X
H19		X	Present	Present	X or SNPs	X
H11		X	Present	Present	X or SNPs	X
H10		X	Present	Present	X or SNPs	X
H9		X	Present	Present	X or SNPs	X
H12		X	Present	Present	X or SNPs	X
H47		X	Present	Present	X or SNPs	X
H45		X	Present	Present	X or SNPs	X
H44		X	Present	Present	X or SNPs	X
H46		X	Present	SNPs	X or SNPs	X

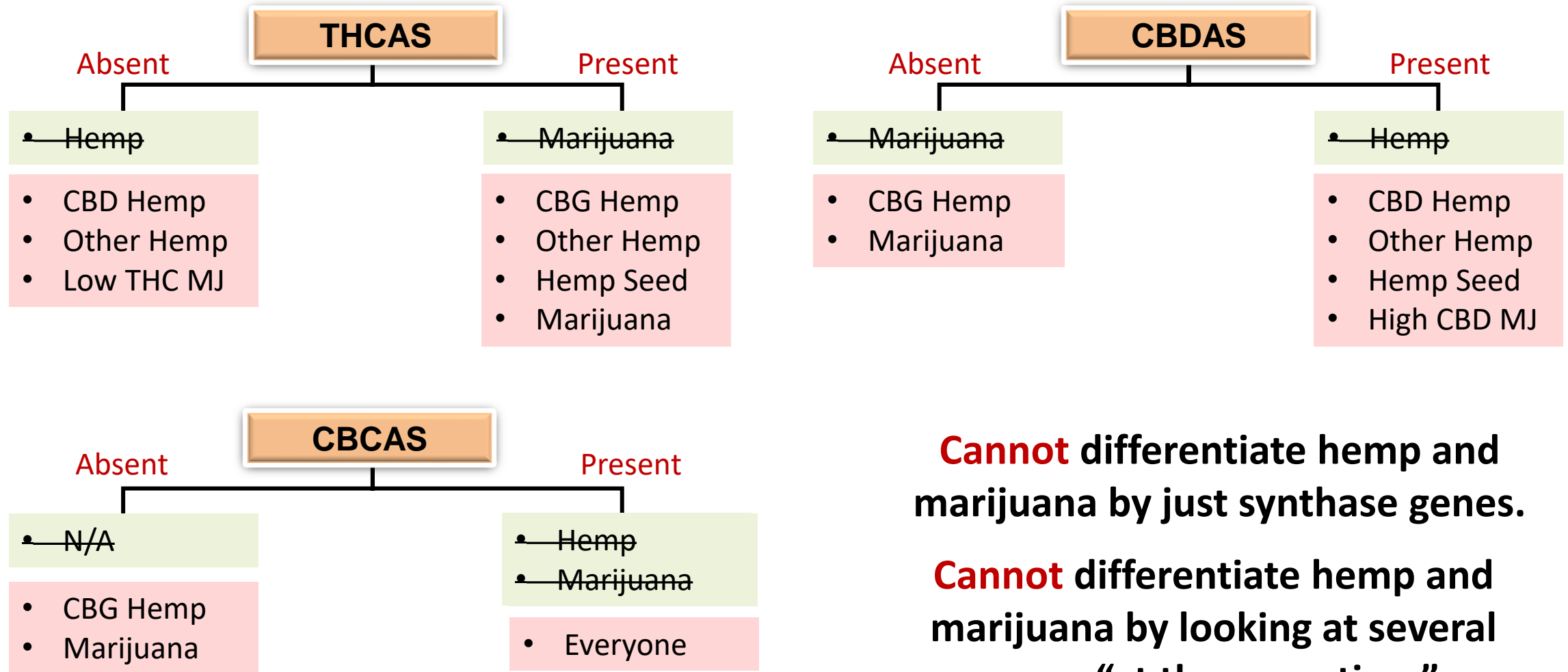
Sample	Strain	THCAS	CBDAS	CBCAS	THCAS-Like	CBDAS-Like
H23	CBG Hemp	SNP 1064G	X	X	Present	Present
CBG30A		SNP 1064G	Incomplete	Present	Present	Present
H22		SNP 1064G	X	X	Present	Present
H21		SNP 1064G	X	X	Present	Present
H41		SNP 1064G	X	X	Present	Present
H42		SNP 1064G	X	X	Present	Present
CBG59A		SNP 1064G	X	Incomplete	Present	Present
H26		SNP 1064G	X	X	Present	Present
CBG22A		SNP 1064G	X	X	Present	Present
CBG40A		SNP 1064G	Incomplete	X	Present	Present
CBG44A		SNP 1064G	Incomplete	X	Present	Present
CBG5A		SNP 1064G	X	X	Present	Present
CBG20B		SNP 1064G	Present	Present	Present	Present

Sample	Strain	THCAS	CBDAS	CBCAS	THCAS-Like	CBDAS-Like
DEA27A	Seized MJ	Present	X	Present	Present	Present
DEA5A		Present	X	Present	Present	Present
DEA36A		Present	X	Present	Present	Present
DEA31A		Present	X	Present	Present	Present
DEA44A		Present	X	Present	Present	Present
DEA42A		Present	X	Present	Present	Present
DEA32A		Present	X	Present	Present	Present
DEA25A		Present	X	Present	Present	Present
DEA60A		Present	X	Present	Present	Present
DEA58A		SNPs	X	Present	Present	Present

Sample	Strain	THCAS	CBDAS	CBCAS	THCAS-Like	CBDAS-Like
H39	CBDV Hemp		Present	Present	X	X
H1-J2	Hemp Seed	SNPs	Present	Present	X or SNPs	Present
H2-5		SNPs	Present	Present	Present	Present
H3-5		SNPs	Present	Present	X or SNPs	Present
H4-4		SNPs	Present	Present	X or SNPs	Present
H36	THCO Hemp	SNPs	Present	Present	Present	Present
H37		X	Present	Present	X	X
H35		X	Present	Present	X	X
H28	Δ8 Hemp	SNPs	Present	Present	SNPs	X
H29		SNPs	Present	Present	SNPs	X
H31		SNPs	Present	Present	SNPs	X
H30		SNPs	Present	Present	X	Present
H34		SNPs	Present	Present	Present	Present
H34		X	Present	Present	X	Present
H32		X	Present	Present	X	Present



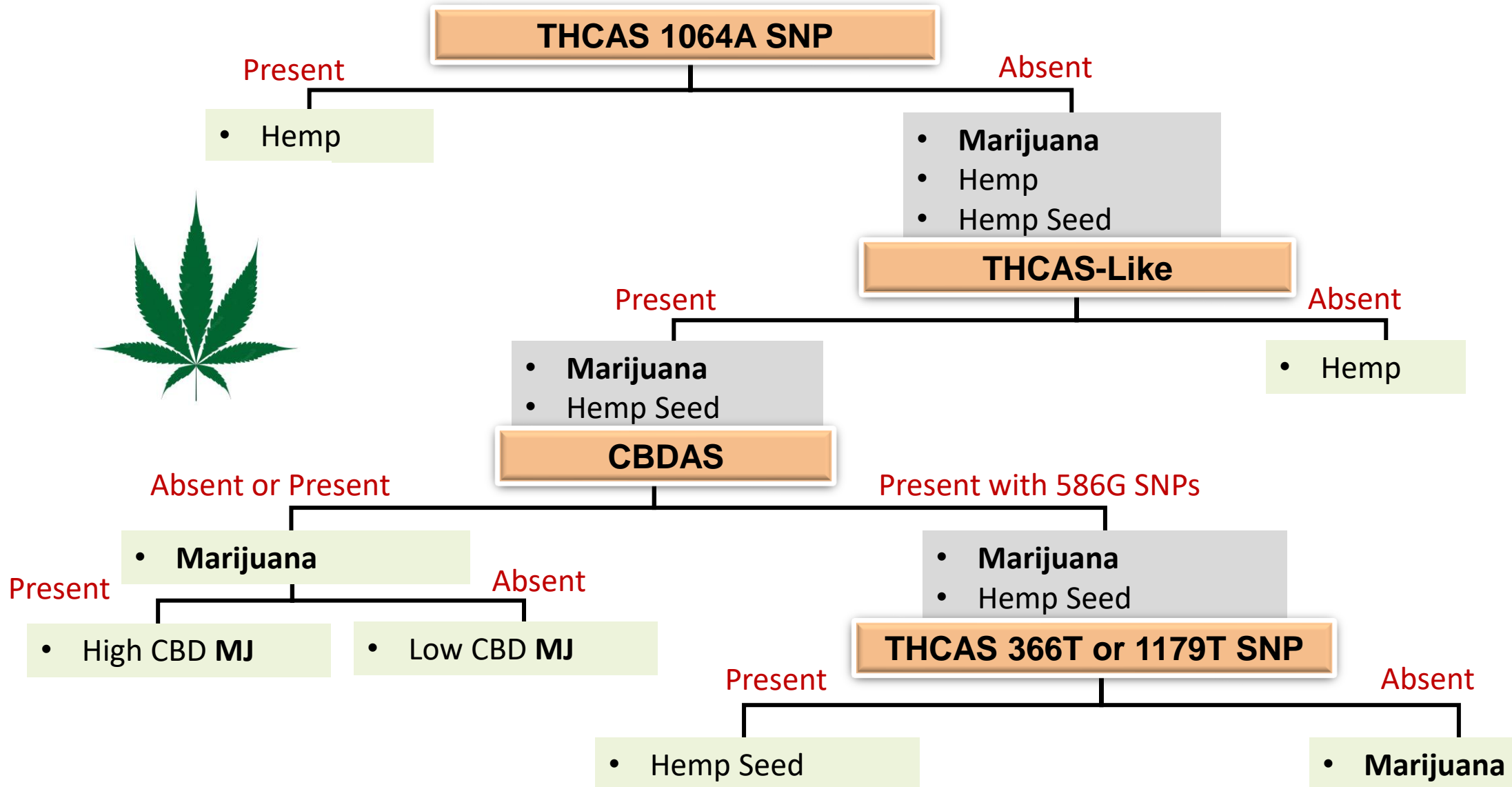
Differentiate Hemp and Marijuana



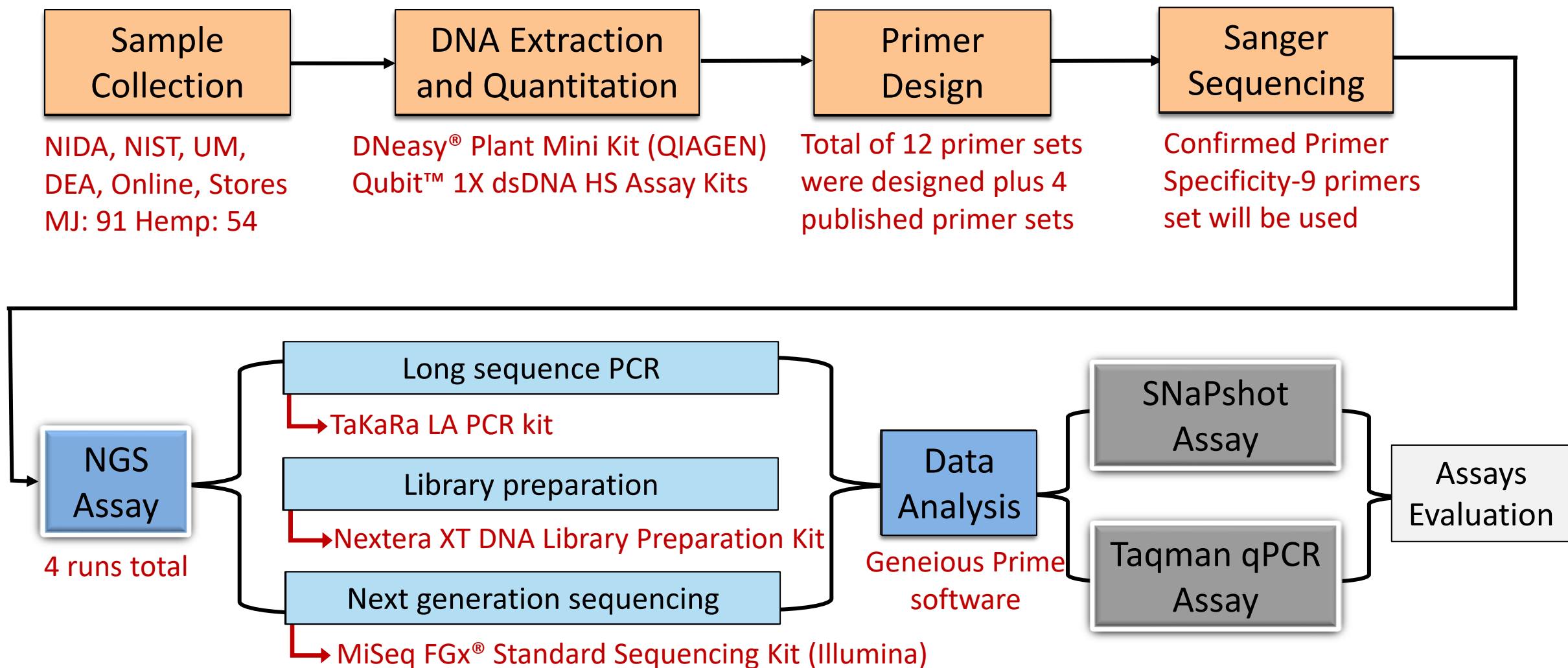
Cannot differentiate hemp and marijuana by just synthase genes.

Cannot differentiate hemp and marijuana by looking at several genes “at the same time”.

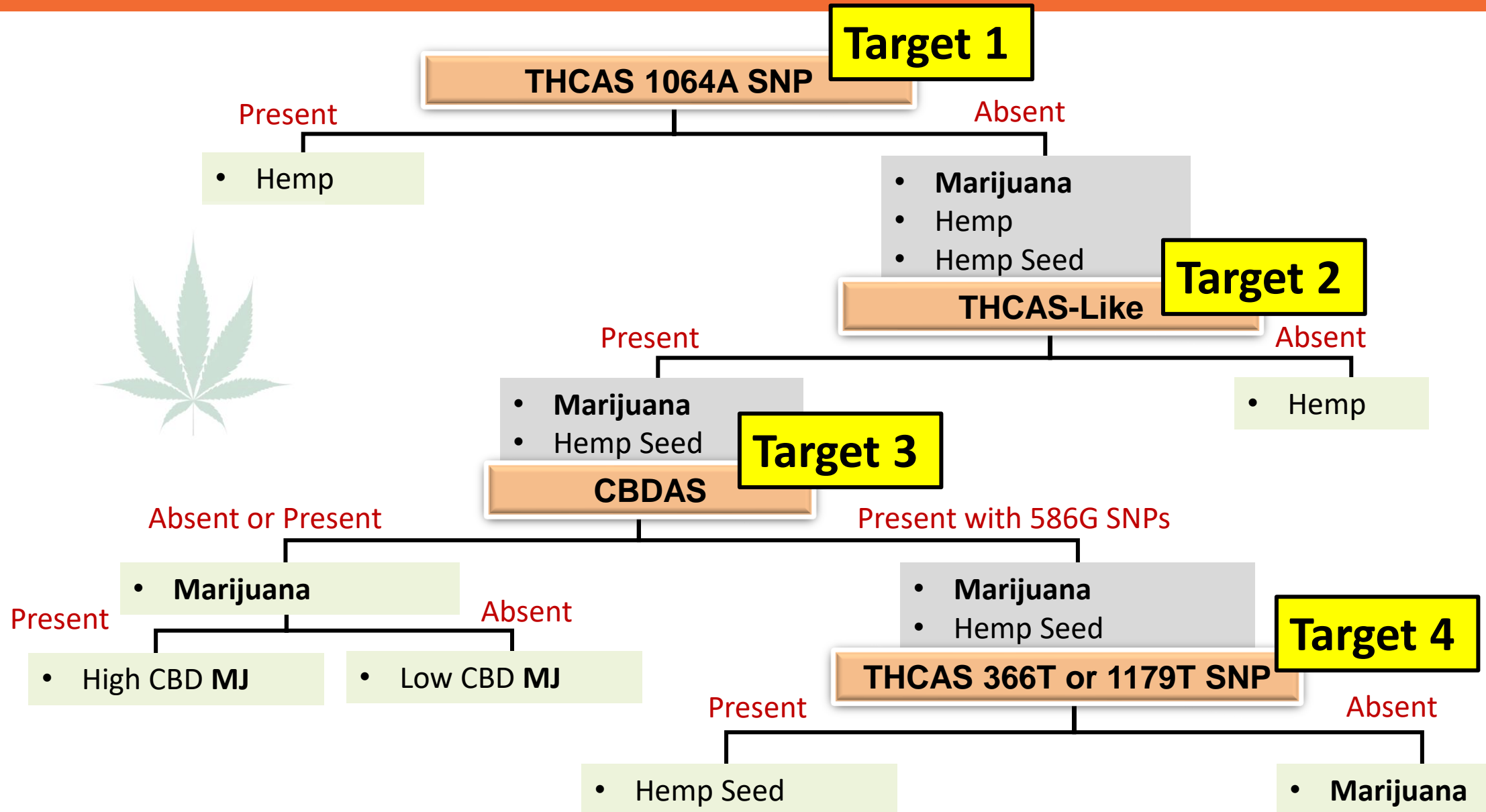
Differentiation Flow Chart



Flow Chart- Next Step



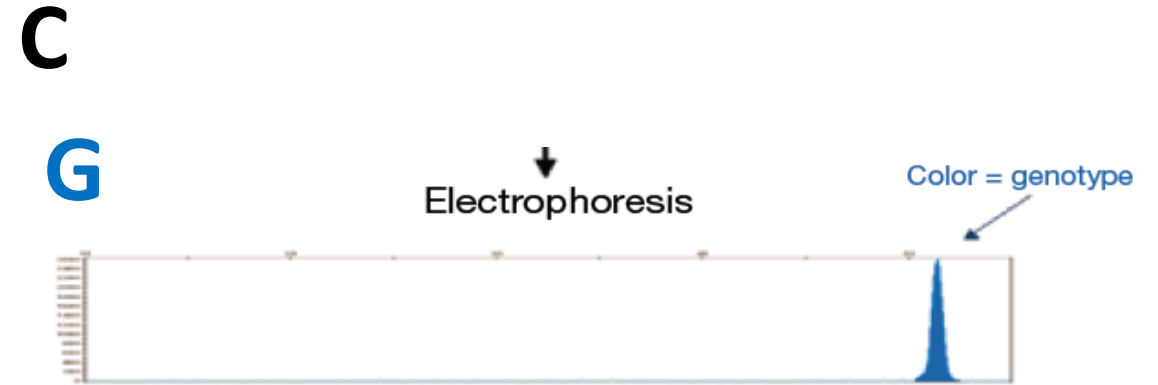
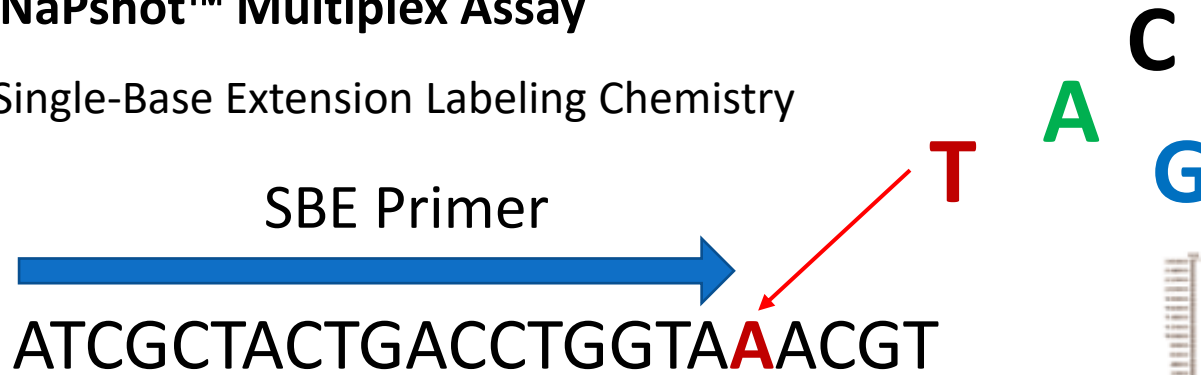
Differentiation Flow Chart-NGS



Development of Fast Genotyping Assay

➤ SNaPshot™ Multiplex Assay

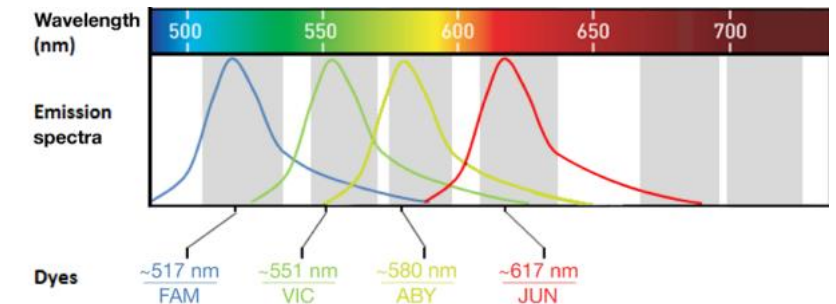
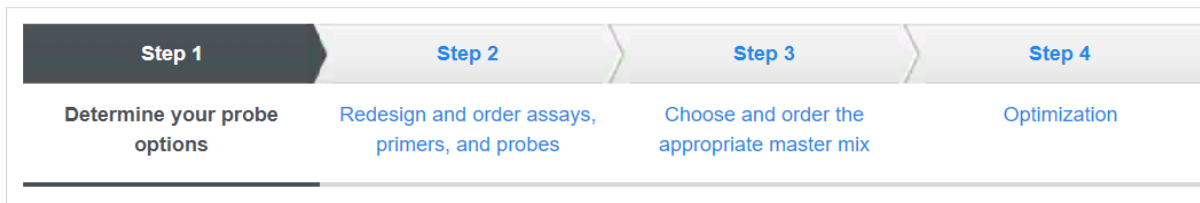
-Single-Base Extension Labeling Chemistry



➤ Custom TaqMan Probes qPCR Assay

Custom QSY probes—QSY probes are best-suited for multiplexed detection of more than two targets and pair with proprietary ABY and JUN dyes that are optimized for Applied Biosystems instruments.

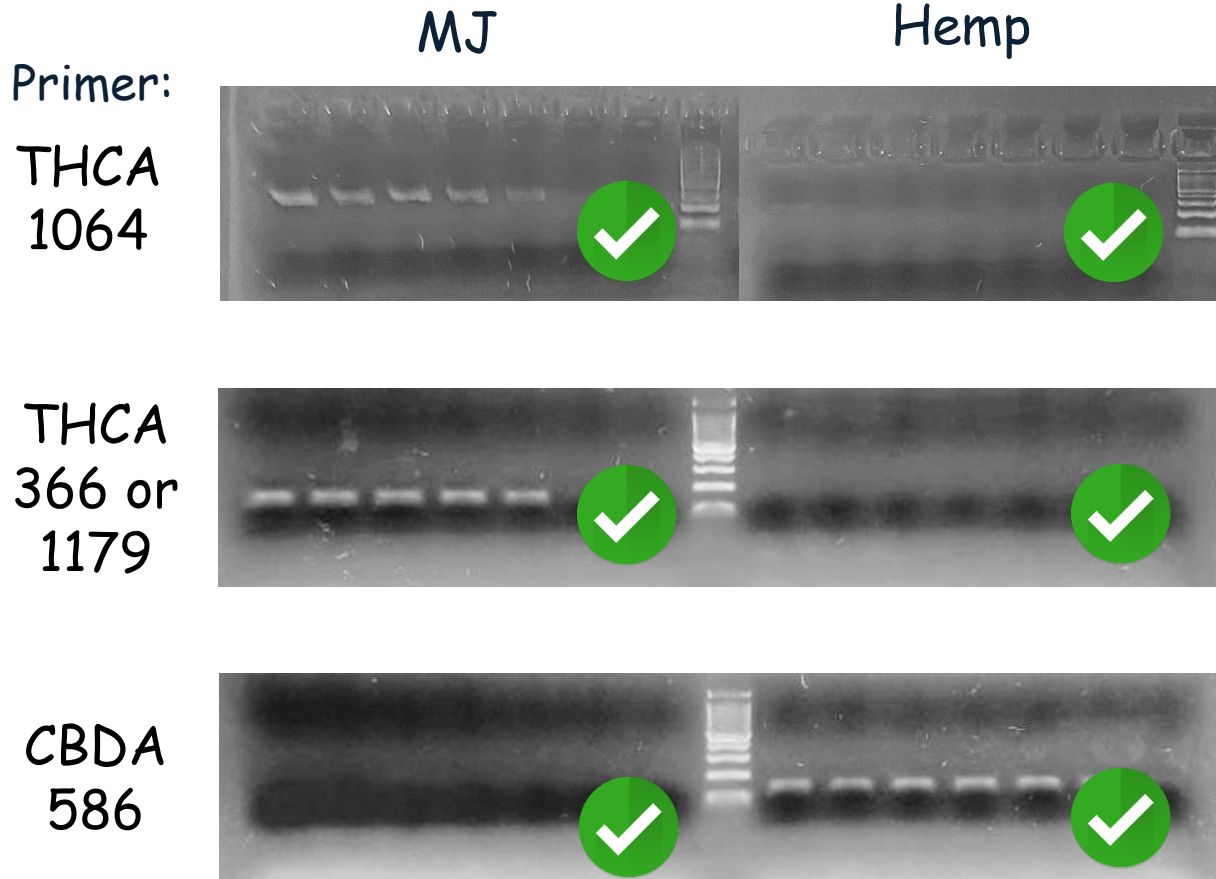
Workflow to develop and optimize your multiplex reaction



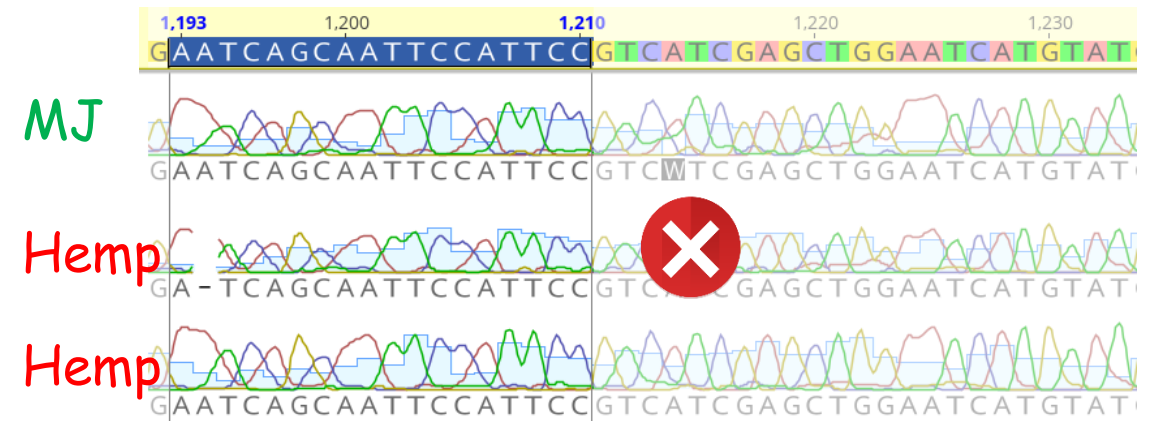
- Gene Expression→4-plex
- SNP Genotyping→2-plex (1 SNP need 2 probes)

First Step-Primer Design

➤ Design 4 PCR primer sets for SNP

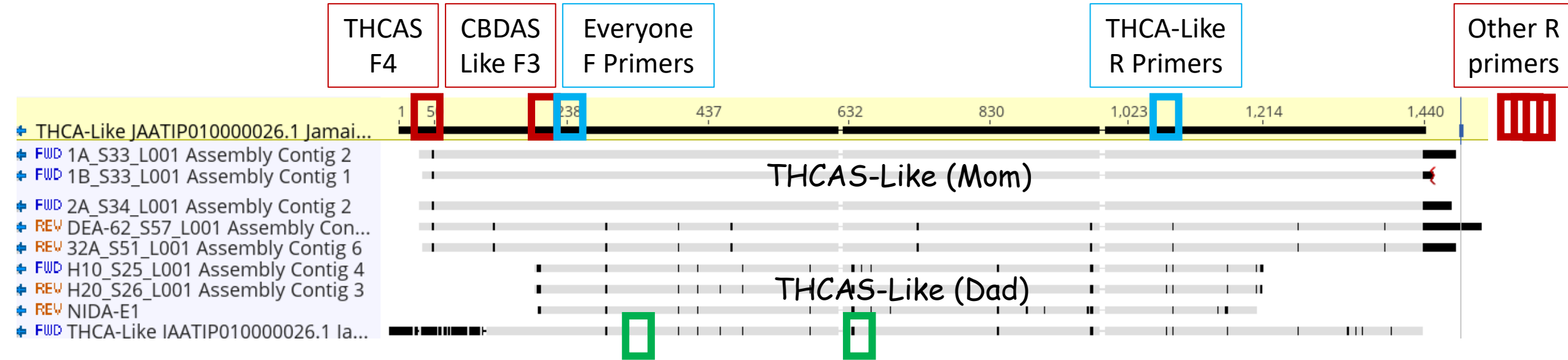


Try different primer sets, try sequencing

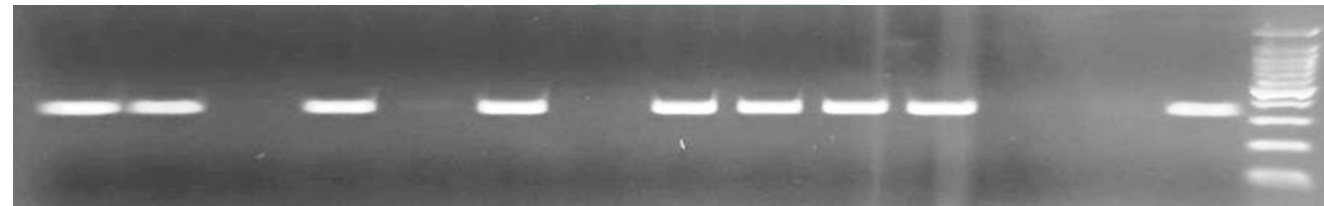


!!Hemp might actually all have THCA-S-Like

NGS vs PCR

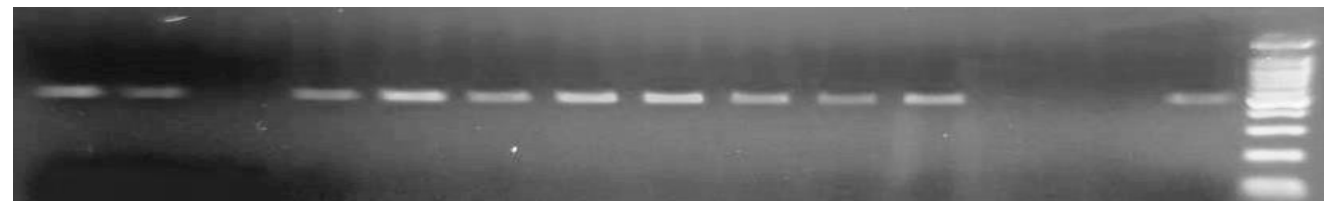


THCAS-Like (M)



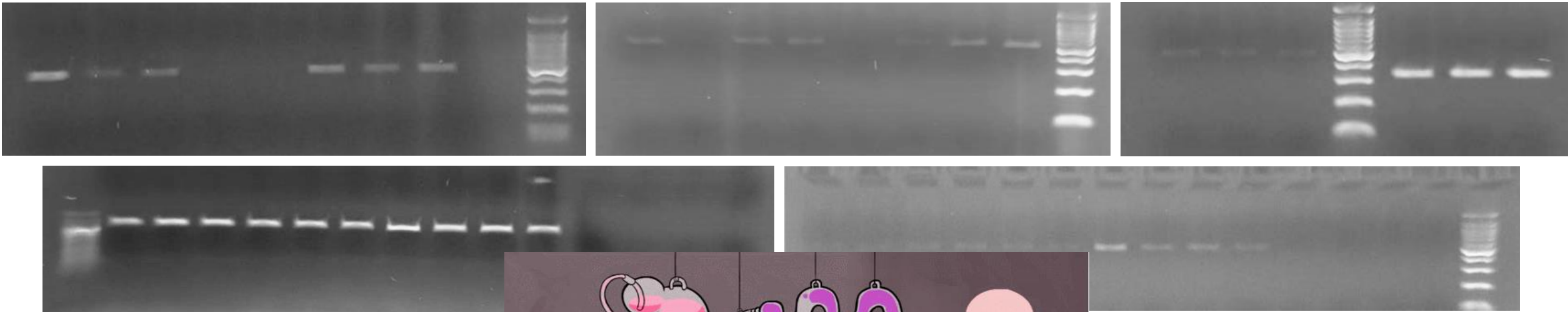
D E F G H29 34 35 36 14B 27B 31B H2 H3 H5

THCAS-Like (D)



THCAS?
THCAS-Like?
Mom or Dad?

Run Gel Run Gel Run Gel...Sequencing Confirmation

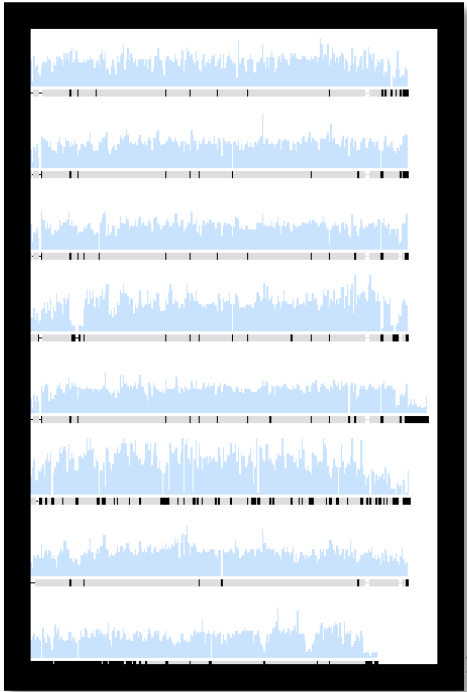


Result	H31	H20	29	61
THCAS Like Mom	V	X	V	X
THCAS Like Dad	V	V	V	X
Results	IH1	IH2	IH3	IH4
THCAS Like Mom	V	V	V	V
THCAS Like Dad	V	V	V	V



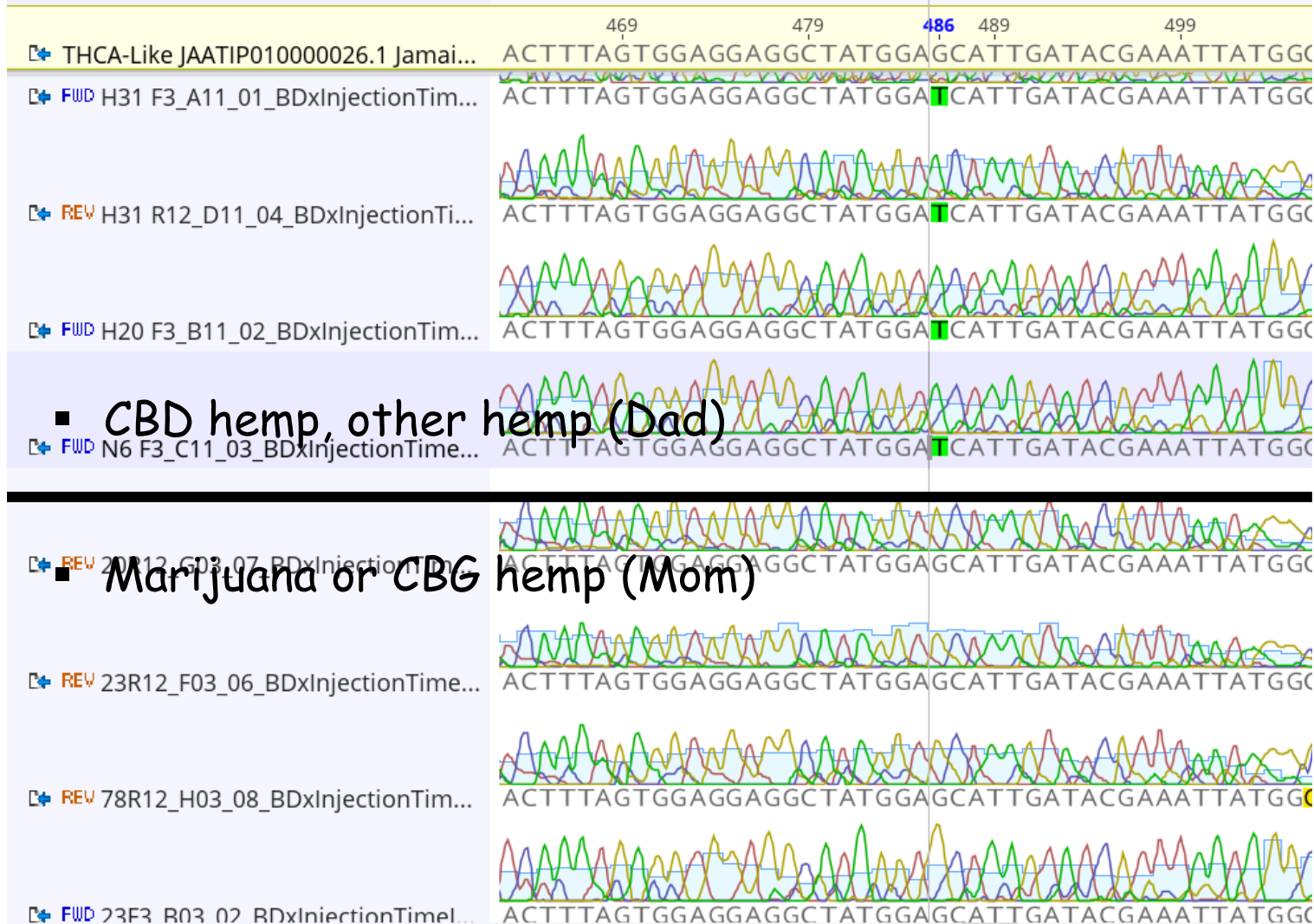
H2-5	H5-5
V	V
V	V
IH13	IH14
X	X
X	X

Result	29B	61B	NIST4	NIST5	NIST11	NIST13	NIST17	NIST10	NIDA H	NIDA I
THCAS Like Mom	V	V	X	X	V	V	F	V	V	V
THCAS Like Dad	V	F	V	V	F	V	V	V	V	V



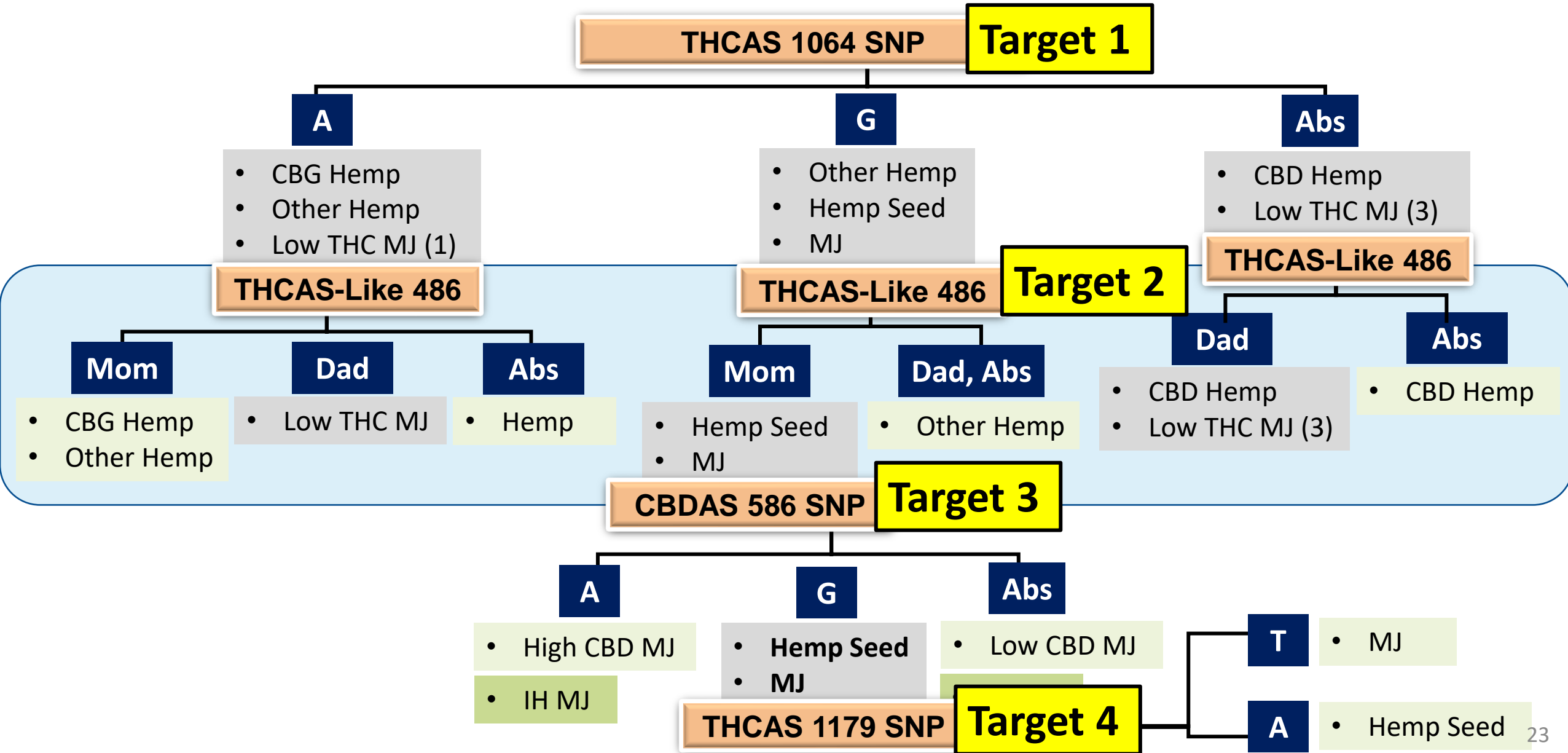
Need to Differentiate the Two Variants of THCAS-Like (Mom and Dad)

- The new primer set will catch Mom in MJ and CBG hemp; catch Dad in CBD hemp

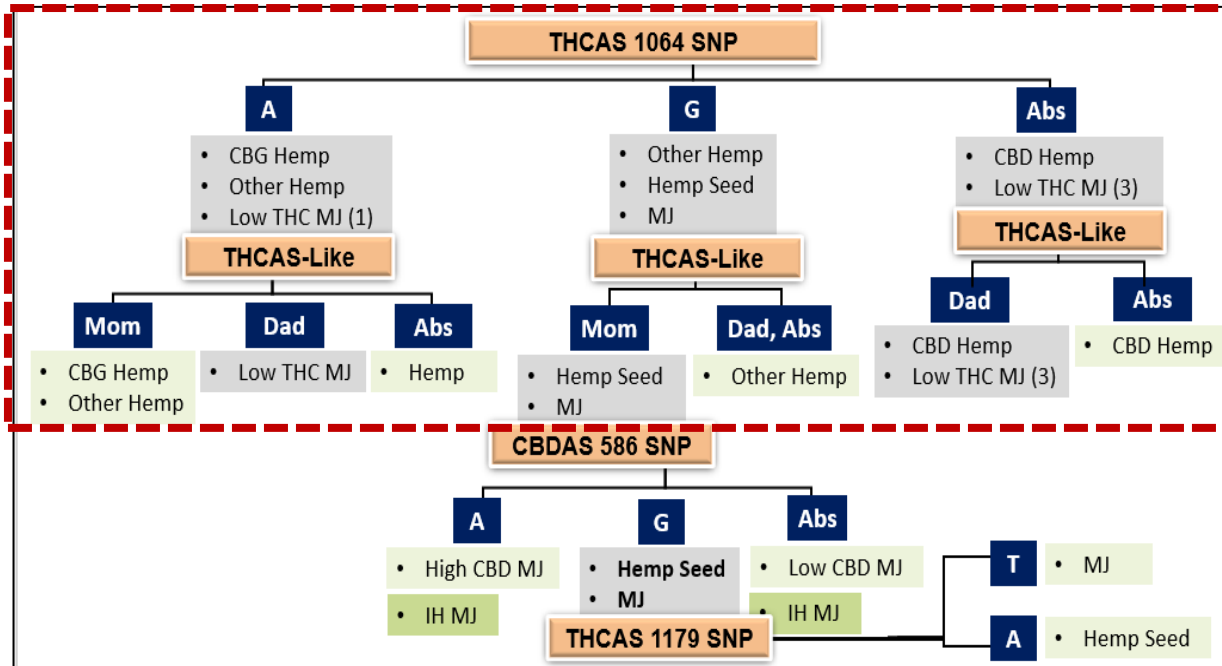


Using THCAS-Like 486 SNP to differentiate Mom and Dad Variants

Differentiation Flow Chart-PCR



Fast Genotyping Assays



Limitation of both:

- Unable to separate some low THC (0.3%-1%) marijuana from hemp

➤ SNaPshot Assay (Can have all targets)

Expectation:

- Separate most Marijuana from all strains of Hemp
- Separate marijuana from hemp seed
- Giving CBD information
- Might be able to estimate THC conc

Main advantage:

- Fast, easier interpretation, provide more info

➤ Tagman qPCR Assay (only allowed two SNPs)

Expectation:

- Separate most MJ from all strains of Hemp
- Might be able to estimate THC conc

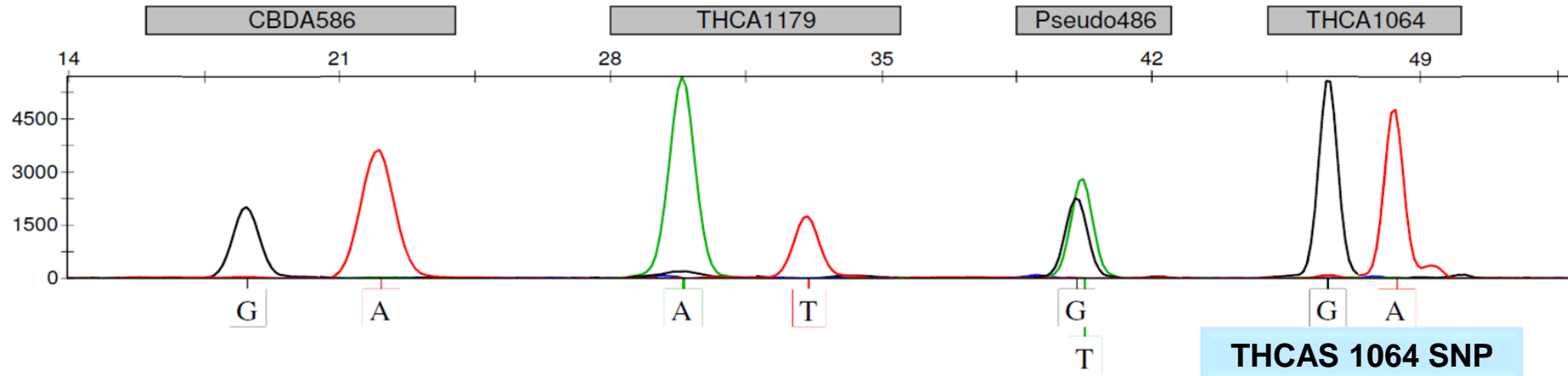
Main advantage:

- Faster than SNaPshot, low cost

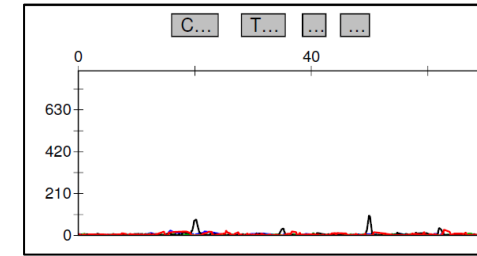
Limitation:

- Unable to separate MJ from hemp seed
- No CBD information

SNaPshot Assay (n=138)



Not Cannabis



CBDAS 586 SNP

THCAS 1179 SNP

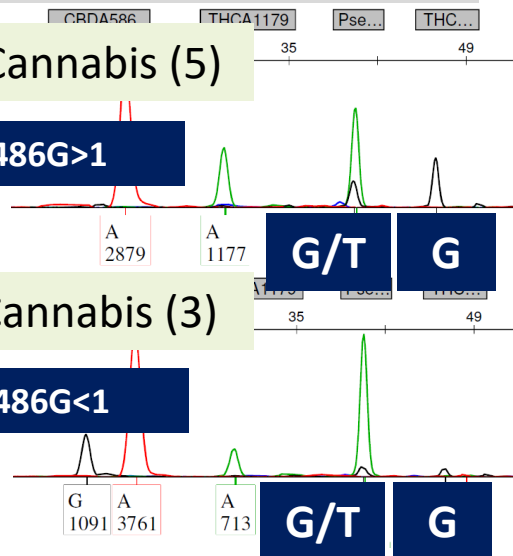
THCAS-Like 486 SNP

THC<1% Cannabis, Hemp Seed (56)

Need some peak height info

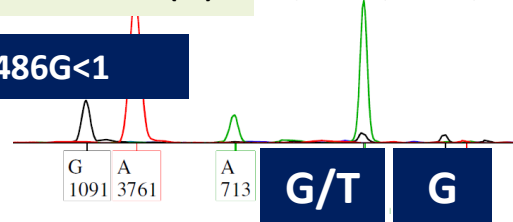
THC>1% Cannabis (5)

1064 G/486G>1



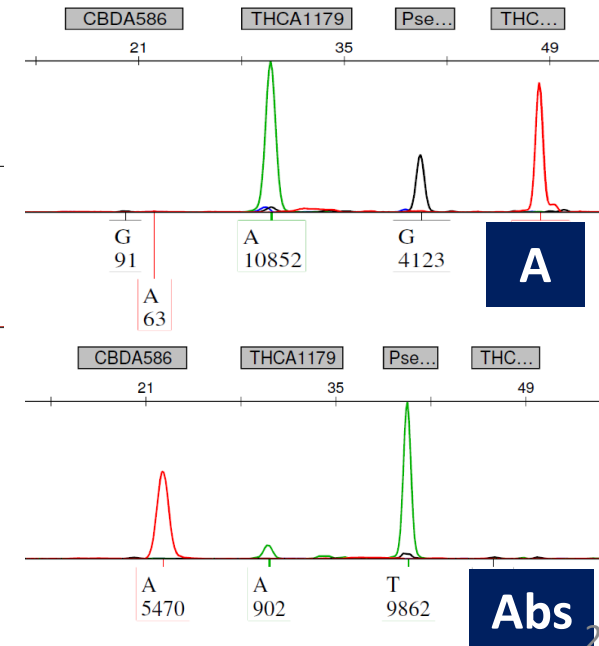
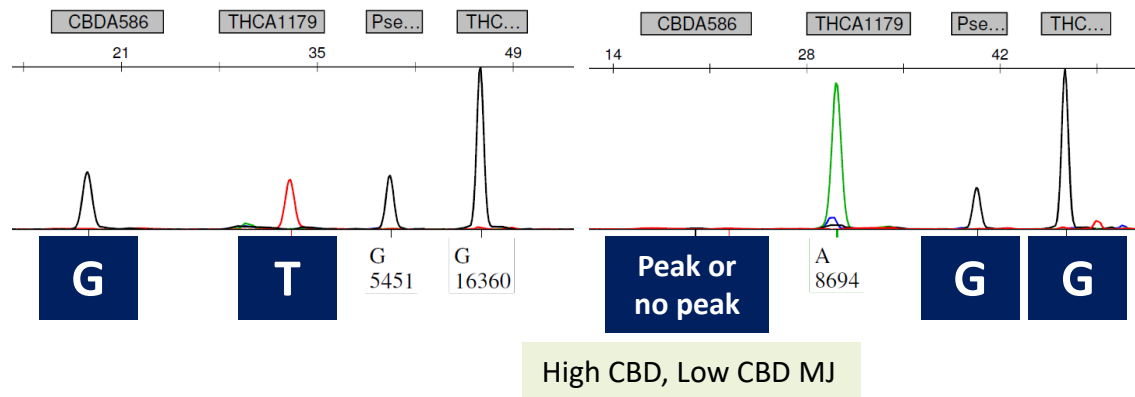
THC<1% Cannabis (3)

1064 G/486G<1



Hemp Seed (2)

Marijuana (69)



Next Step

Tagman qPCR Assay



Next Step-Assay Validation

Blind testing Cannabis sample to validate the application of new genotyping assay

Strain	Number	Source	Chemistry Data
US-Mexico Marijuana	60	CBP LSSD	Not Available
Chile Marijuana	20	PDI	Not Available
Chile Medical Marijuana	4	PDI	Not Available
Commercial Hemp	Need More!	Shops	Hopefully Available
Seized Marijuana	Need More!		Hopefully Available
Reference Sample	Need More!		Hopefully Available
Hemp Seed	Need More!	Shops	Hopefully Available
Other Type of Sample	Needed		Hopefully Available
Total	150-200		

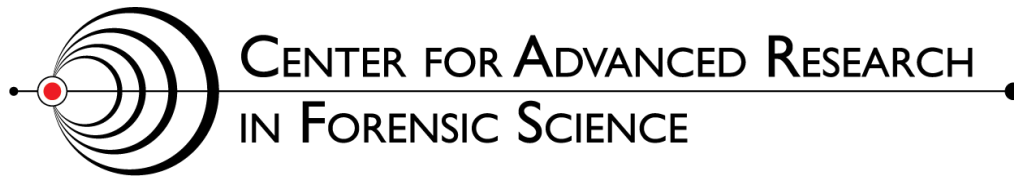


Summary

- Demonstrated the usage of the custom targeted NGS panel that serves as a comprehensive analysis of the three synthase genes and their pseudogenes to discover the relationship between chemotype and gene copies.
- With wider variety of sample types, hemp and marijuana cannot be differentiated look at only THCAS and CBDAS gene.
- Potential comprehensive workflow to differentiate cannabis crop types with not only synthase genes but the pseudogene.
- Develop a rapid SNaPshot assay for the differentiation of hemp and marijuana (validation need in future study).

Acknowledgements

- Center for Advance Research in Forensic Science (CARFS)
- National Institute on Drug Abuse (NIDA) – Drug Supply Program
- National Institute of Standards and Technology (NIST)
- University of Mississippi





THANK YOU FOR YOUR ATTENTION



[Jessica \(Ya-Chih\) Cheng](#)

yxc059@shsu.edu

