The development of fast genotyping assays targeting cannabinoid synthase genes for the differentiation of hemp and marijuana

**Abstracts (<250 words)**

The legalization of hemp cultivation in the United States has raised the need for reliable methods to distinguish between legal hemp and illegal marijuana. Genetic analysis has emerged as a powerful tool, surpassing traditional chemical methods in specific aspects, such as analyzing trace amounts, aged samples, and different parts of the sample. Genetic differences in cannabinoid synthase genes offer promise for precise crop type determination, particularly focusing on genes like tetrahydrocannabinolic acid synthase (THCAS), cannabidiolic acid synthase (CBDAS), and cannabichromenic acid synthase (CBCAS). However, previous research faced several challenges in developing discriminatory genetic markers, including limited sample sizes, high similarity between the synthase genes, and the present of pseudo synthase genes. A comprehensive study using Next-Generation Sequencing (NGS) introduced a differentiation flowchart based on THCAS, CBDAS, and THCAS pseudogenes. To bridge the gap between the NGS platform and the practical requirements of crime labs, two rapid genotyping assays were developed: a CE-based SNaPshotTM assay and a TaqManTM real-time PCR assay. While the SNaPshotTM assay effectively differentiated various hemp and marijuana types, differentiation was limited with marijuana samples containing THC% close to the 0.3% legal threshold (0.3-1%). The TaqMan q-PCR SNP genotyping assay provided quicker results, making it an efficient choice for crime labs. However, this method had challenges with differentiating edible hemp seed samples, and it did not provide additional CBD information. The study also highlighted the influence of two variants of one THCAS pseudogene on chemotype determination, emphasizing the necessity for precise genetic analysis for accurate categorization of cannabis varieties.

**KEYWORDS:** Marijuana, Hemp, Cannabinoid synthase, Pseudogene, SNaPshotTM assay, TaqManTM qPCR assay

**HIGHLIGHTS (<100 characters, 3-5 bullet points)**

* New differentiation flowchart with novel synthase markers for hemp and marijuana.
* A SNaPshot™ assay using novel synthase markers was developed and validated with 130 samples.
* A TaqMan™ genotyping assay using novel synthase markers was developed and validated with 130 samples.
* Examined the THCAS pseudogene's potential role in cannabis chemotype.

**Introduction**

Hemp and marijuana, both derived from *Cannabis sativa* L*.* (*C. sativa*), are distinguished by their legal status depending on the amount of delta-9-tetrahydrocannabinol (Δ9-THC). With hemp legally defined as *C. sativa* containing less than 0.3% Δ9-THC by dry weight under the 2018 Farm Bill, the need for efficient analytical methods to differentiate it from federally prohibited marijuana is vital in the United States (1, 2). DNA-based forensic botany techniques have proven powerful for cannabis analysis, surpassing traditional chemical methods in specific aspects. Genetic tools offer promise in overcoming limitations seen in chemical profiling, including variations in cannabinoid content due to environmental factors, storage conditions, harvest time, and plant parts (3, 4). Moreover, genetic analysis utilizes smaller tissue samples compared to larger sample sizes required by chemical analyses (5). Prior investigations have highlighted the genetic distinction between hemp and marijuana cultivars (6, 7), leading to the proposal of various genetic tools for the identification and individualization of *C. sativa* samples (8-12). Among these strategies, a focus has been placed on utilizing genes associated with cannabinoid production enzymes to differentiate between marijuana and hemp (10, 13, 14). Specifically, tetrahydrocannabinolic acid synthase (THCAS), cannabidiolic acid synthase (CBDAS), and cannabichromenic acid synthase (CBCAS) serve as pivotal catalysts in the conversion of the precursor molecule cannabigerolic acid (CBGA) into three distinct cannabinoid acids THCA, CBDA, and CBCA (15, 16). These cannabinoid acids subsequently undergo decarboxylation upon exposure to heat or combustion, yielding the bioactive compounds tetrahydrocannabinoid (THC), cannabidiol (CBD), and cannabichromene (CBC) (17). Consequently, the polymorphisms in the genetic code of these enzymes were assumed to play a role in delineating the chemotype or chemical composition characterizing a particular cultivar. However, previous attempts to identify and validate genetic markers on these synthase genes for crop type identification have encountered limitations.

Firstly, previous research has been limited by small sample sizes and restricted sample varieties (10, 18, 19). A wider sample variety and increase in sample numbers are needed for a more comprehensive approach to enhance the reliability and applicability of genetic markers in cannabis studies. Secondly, the cannabinoid synthase genes are reported to be highly similar, which means primers with high specificity need to be more carefully designed (20-22). Furthermore, cannabis genomic research has unveiled the presence of highly repetitive cannabinoid synthase genes, including numerous pseudogenes with incomplete coding sequences or internal stop codons. These pseudogenes can exhibit nucleotide identities of 91% to 95% when compared to functional synthase genes, posing a potential risk of unintended pseudogene amplification in genetic testing methodologies for cannabis (22-25). To address the aforementioned challenges and risk, Cheng et al. conducted an extensive analysis of widely recognized synthase genes (THCAS, CBDAS, CBCAS) and synthase-like genes (pseudogenes) across a diverse sample set using next-generation sequencing (NGS) (26). The sample set included reference cannabis samples (n=30), a wide range of marijuana samples from various states (n=60), and seven distinct hemp types categorized by major cannabinoid content (n=57). The study proposed a differentiation flowchart based on the sequential analysis of gene patterns and specific single nucleotide polymorphisms (SNPs) of THCAS, CBDAS, and THCAS pseudogenes, successfully distinguishing marijuana samples from various hemp strains, including CBD hemp, CBG hemp, Δ8-THC hemp, Δ10-THC hemp, THCO hemp, CBDV hemp, and hemp seeds (26).

The NGS platform and existing differentiation flowchart (26), while effective, may not be cost-effective for analyzing case work samples in crime laboratories. Therefore, this study aimed to develop two rapid genotyping assays: a capillary electrophoresis (CE)-based SNaPshotTM assay and a polymerase chain reaction (PCR)-based TaqMan real-time PCR assay, targeting the SNPs and gene expression markers identified in the differentiation flowchart. Compared to the traditional forensic laboratory Short Tandem Repeat (STR) workflow (27), the SNaPshotTM assay requires no additional instrumentation and the q-PCR assay yields result with only extraction and quantitation instrument. The SNaPshotTM assay, utilizing Life Technologies’ SNaPshotTM kit (ABI PRISM® SNaPshotTM Multiplex Kit (28)), employs a single-base extension (SBE) reaction and detects the resulting dye-labeled SBE products through CE, providing simplified multiplex SNPs typing and cost-effective workflow (29). Similarly, the real-time PCR assay offers high efficiency and simplicity, gaining prominence in forensic applications such as species and body fluid identification (30-32). Multiplex real-time PCR, utilizing TaqMan chemistry, provides high sensitivity in duplex reactions and serves as a valuable alternative to traditional PCR methods relying on sequence-specific probes (33, 34). In this study, both assays were designed using the differentiation flowchart and markers proposed in the previous study (26). Diverse cannabis samples, including samples from various seized and medical marijuana sources outside the United States, will be genotyped using these assays to ensure their reliability and applicability. The study will also assess the advantages and limitations of these assays, catering to crime laboratories with specific needs and available instrumentation, thereby assisting in the differentiation of hemp and marijuana.

**Materials and methods**

**Sample collection, extraction, and quantification**

*C. sativa* with known cannabinoid concentrations (Total THC, total CBD, and total CBC) provided by National Institute on Drug Abuse (NIDA), National Institute of Standards and Technology (NIST), and the University of Kentucky served as reference materials for marijuana (n = 27) and hemp (n = 3). A total of forty-five commercial hemp samples were collected from diverse online vendors within the United States. Nineteen samples of CBD hemp flowers (n = 19) were procured, with origins spanning The Original Hemp Buds (OR or NY), CBD Hemp Direct situated (Las Vegas, NV), Industrial Hemp Farms (Denver, CO), and VAST Wellness Solutions (Dallas, TX). For CBG hemp flowers (n = 8), our sources included CBD Hemp Direct (Las Vegas, NV), Industrial Hemp Farms (Denver, CO), ATLRx Premium CBD Flower (Atlanta, GA), and VAST Wellness Solutions (Dallas, TX). Additionally, an array of other hemp strains, including Δ8-THC hemp (n = 8), Δ10-THC hemp (n = 1), THCO hemp (n = 3), and CBDV hemp (n = 1), were acquired from CBD Supply Marylandthe (Baltimore, MD), Green Unicorn Farm (Fairfax, CA), as well as ATLRx Premium CBD flower (Atlanta, GA). Hemp seed samples (n = 5) were sourced from local grocery stores. Extracts from marijuana seized by the Drug Enforcement Administration (DEA) were generously provided by the University of Mississippi (n = 24). The cannabinoid concentration of the DEA seized samples is also provided by the University of Mississippi. Furthermore, a collection of US-Mexico marijuana samples (n = 14), including flower, stem, and seed specimens, were extracted from previously processed cases at the US Customs and Border Protection (CBP). DNA extracts of marijuana from southern Chile (n = 13) were obtained from the Policia de Investigaciones. Three medical marijuana extracts from Chile (n = 3) were contributed by our collaborators in Chile. Lately, one strain of *Humulus lupulus* (Hops) was also included in the validation (n=1). The DNA extraction methods varied (Table 7.1.), with in-house extraction performed using the DNeasy® Plant Mini kit (QIAGEN, Hilden, Germany) following the manufacturer's protocols (35). Genomic DNA quantification was conducted using the Qubit dsDNA HS Assay Kit System (Thermo Fisher Scientific, South San Francisco, CA, USA) as per the manufacturers’ instructions (36). The StepOne™ Real-Time PCR System (Thermo Fisher Scientific) was also employed for some extracts (37). The cannabinoid concentrations of the commercial hemp and reference plant materials were obtained from the provider's certificate of analysis. For the DEA seized sample, cannabinoid concentrations were provided by the University of Mississippi. Additional details about the samples and strains can be found in Supplementary Table S1.

**Marker confirmation and custom PCR primer designed for genotyping assays**

A differentiation flowchart was proposed in previous study involving THCAS, CBDAS, and one pseudo THCAS gene (26). Specific genetic markers, namely SNPs at position 366 and 1064 bp of the THCAS gene (AB212837.1), the SNPs at position 586 of the CBDAS gene (B292682.1), and the expression of the THCAS pseudogene (THCAS-Like, GCA\_012923435.1), were used for the development of the rapid genotyping assays. For the amplification of SNP regions within the THCAS and CBDAS genes, primers were designed using Primer3web version 4.1.0 (https://primer3.ut.ee/). The primer sequences were designed to be approximately 20 bp in length with amplification products ranging from 100 to 300 bp. To detect the presence or absence of the THCAS-Like gene, a comprehensive alignment of 123 reported synthase genes and pseudogene sequences was performed (Supplementary Table S2, (26). Regions exhibiting high specificity to the THCAS-Like JAATIP010000026.1 sequence, and with a sequence length of less than 300 bp, were chosen as the binding sites for the forward and reverse primers of the THCAS-Like gene. No interaction between and within the primer sets were confirmed with Multiple Primer Analyzer (38). The optimal annealing temperature for the multiplex PCR was determined through gradient PCR using Type-it® Microsatellite PCR Master Mix (QIAGEN), following the manufacturer's protocol (39). Each PCR reaction mixture contained 6.25 μL of Type-it Microsatellite PCR Master Mix, 1.25 μL of 5X Q solution, 1.25 μL of the primer mix, 2 μL of diluted template DNA (1 ng/μL), and 1.75 μL of water. The final PCR conditions consisted of an initial 5-minute enzyme activation step at 95 °C, followed by 30 cycles of denaturation at 95 °C for 30 seconds, annealing at 58 °C for 90 seconds, and extension at 72 °C for 60 seconds. The final extension step was performed at 60 °C for 30 minutes. To confirm the specificity of the primers, experimental validation was conducted using four cannabis extracts, including one CBG hemp (H23), two CBD hemp (H25, H45), and one marijuana strain (NIST-2022-20). Sanger sequencing was performed to confirm SNPs of the four tested extracts. Sequencing reactions were carried out using the BigDye Terminator™ v3.1 cycle sequencing kit (Thermo Fisher Scientific) in accordance with the manufacturer's guidelines (40). The sequencing products were purified using the BigDye XTerminator™ Purification Kit (Thermo Fisher Scientific) and analyzed on a 3500 Genetic Analyzer (Thermo Fisher Scientific) with a 50 cm capillary and POP-7™ polymer. Sequence data were comprehensively analyzed using Geneious Prime Software (Dotmatics, Boston, MA).

**SNaPshot™ assay development and genotyping**

PCR reactions consisted of 6.25 µL of Type-it® Microsatellite PCR Master Mix (QIAGEN), 1.25 µL of primer mix, 1.25 µL of 5X Q-solution (QIAGEN), 1.75 μL of water, and 2 µL of DNA sample (2 ng). Cycling conditions consisted of a 5 min enzyme activation at 95 °C; 30 cycles of 30 s at 95 °C, 90 s at 58 °C, and 60 s at 72°C; and a 30 min final extension at 60 °C. PCR cleanup was achieved using 2 µL of ExoSAP-IT™ (Thermo Fisher Scientific) mixed with 5 µL of PCR product. Samples were incubated at 37 °C for 15 minutes, followed by an enzyme inactivation at 80 °C for 15 minutes.

Single base extension (SBE) reactions were executed using half-reactions of the SNaPshotTM Multiplex Kit (Thermo Fisher Scientific), strictly adhering to the manufacturer’s established protocol (28). To facilitate precise SNP detection, SBE primer sequences were designed, spanning 15-25 base pairs before (forward) or after (reverse) the targeted SNPs. Selection criteria for the SBE primers included minimal interactions both between and within the primers, a parameter evaluated using the Multiple Primer Analyzer tool (38). Following the thermal cycling steps as per the manufacturer’s guidelines, a cleanup procedure was initiated. This involved the addition of 1 µL CIAP (1 U/µL, Promega Corporation) and an incubation step at 37 °C for 15 minutes, succeeded by enzyme inactivation at 80 °C for 15 minutes. Subsequently, capillary electrophoresis (CE) was conducted using a 3500 Genetic Analyzer (Thermo Fisher Scientific). The run parameters were configured as follows: oven set at 60 °C; pre-run phase at 15kV for 180 seconds; injection at 1.6 kV for 8 seconds; main run at 15 kV for 560 seconds; capillary length of 36 cm; POP-4™ polymer; and dye set E5. Data obtained from the CE runs were analyzed utilizing GeneMapper ID v.5 software (Thermo Fisher Scientific), employing an analytical threshold of 200 RFU for accurate interpretation.

**TaqMan real-time PCR SNP genotyping assay design**

The TaqMan® Custom SNP Genotyping Assays utilized in this study were capable of detecting only two SNPs simultaneously. Therefore, the assay focused solely on the detection of the 1064 SNP on THCAS and the presence of THCAS pseudogene. In this assay, TaqMan QSY probes (Thermo Fisher Scientific) were employed due to their unique feature of accommodating four 5ʹ reporter dye options, namely Applied Biosystems™ FAM™, VIC™, ABY™, and JUN™ dyes (41). This versatility enabled the simultaneous detection of the two target SNPs. The probes were designed to have a melting temperature (Tm) approximately 10 °C higher than that of the PCR primers, ensuring optimal binding specificity to the target sequences (42). The final probe was selected to minimize interactions both between and within the probes, a selection process facilitated by the Multiple Primer Analyzer tool. To enhance the multiplexing capabilities, the TaqPath™ ProAmp™ Multiplex Master Mix with MUSTANG PURPLE™ (Thermo Fisher Scientific) was chosen (43). This specific master mix was selected due to its unique ability to incorporate four probes simultaneously within the reaction. Custom primers, designed with the previously established sequences, were also purchased from Thermo Fisher Scientific to ensure efficient amplification within the master mix.

Real-time PCR was conducted using QuantStudio™ 5 Real-Time PCR System (Thermo Fisher Scientific). Reactions consisted of 6 µL of TaqPath™ ProAmp™ Multiplex Master Mix, 0.6 of primer mix, 0.4 μL of water, and 5 µL of DNA sample (1 ng). Cycling conditions starting with 30 s at 65 °C and 5 min at 95 °C; 40 cycles of 15 s at 95 °C and 60 s at 60°C; end with a 30 s post-read at 60 °C. Data were analyzed utilizing Thermo Fisher Cloud Genotyping Application (Thermo Fisher Scientific).

**Results and Discussion**

**Markers and PCR primers confirmation**

Primer sets targeted SNPs at positions 366 and 1064 bp within the THCAS gene, and SNPs at position 586 within the CBDAS gene were successfully designed (Table 1). Specificity was assessed by testing against four extracts, and the results confirmed the expected outcomes. However, the initial assessment of THCAS pseudogene expression yielded variable outcomes. Despite attempting three different primer sets (THCAS-Like 1-3), combined with gel electrophoresis and Sanger sequencing confirmation, the obtained results remained inconsistent with the expected outcome (Table 1). The prior investigation, employing NGS, indicated the absence of the THCAS pseudogene (THCAS-Like) in certain hemp strains (see Chapter VI) while the current study revealed the presence of the pseudogene in samples where it had gone undetected by NGS. Consequently, a more comprehensive evaluation of THCAS pseudogene expression was undertaken. The THCAS-Like sequences examined in this study were observed originally in the *C. sativa* cultivar Jamaican Lion (44). The Jamaican Lion possesses both female parent (GCA\_012923435.1) and male parent (GCA\_013030025.1), and the THCAS-Like sequence on chromosome JAATIP010000026.1 exhibits several SNP differences between the parents (Supplementary Figure S1). Upon closer analysis of the NGS results, along with the primers used in the previous study (26) and our new primer designs, it became evident that the THCAS-Like sequence from the female plant aligns with the THCAS forward primer (THCAS F4), whereas the THCAS-Like sequence from the male plant aligns with the CBDAS-Like forward primer (CBDAS+Like F3). Consequently, we paired these two forward primers with new PCR reverse primers (THCAS-Like R2, THCAS-Like R3) to investigate the THCAS-Like gene expression pattern. Sanger sequencing confirmed these sequences, revealing that THCAS-Like primer set 4 (Table 1, THCAS F4 with THCAS-Like R2) captured both THCAS and THCAS-Like genes, while THCAS-Like primer set 5 (Table 1, CBDAS+Like F3 with THCAS-Like R3) specifically targeted THCAS-Like genes. Interestingly, Sanger sequencing results indicated that the THCAS-Like gene in most hemp strains corresponds to the sequence from the male plant, whereas in marijuana and hemp seeds, it aligns with the THCAS-Like sequence from the female plant (Supplementary Figure S1). Consequently, to adhere to the differentiation flowchart for distinguishing between hemp and marijuana, it became crucial to differentiate between these two variants of the THCAS-Like gene. THCAS-Like primer set 5 was selected due to its higher specificity compared to primer set 4. Furthermore, a single nucleotide polymorphism (SNP) at position 486 on the THCAS-Like sequence was identified as the key differentiator between the female and male variants (Supplementary Figure S2).

**SNaPshot™ assay development and genotyping**

**Assay development**

Initially, a four-plex SNaPshotTM assay was developed for the genotyping of four SNP loci, which included the 1064 and 366 SNPs on the THCAS gene (THCA1064, THCA366), the 586 SNP on the CBDAS gene (CBDA586), and the 486 SNP on the THCAS-Like gene (Psedo486). However, it was consistently observed that the genotype of the 366A SNP appeared prominently in every sample, even in those containing the 366T SNP. Subsequent investigation through Sanger sequencing revealed that the close proximity of the 366 THCAS SNP and the 486 Pseudo-THCAS SNP, along with the sequence similarity between the THCAS and THCAS-Like genes, resulted in the 366 SNP probe inadvertently capturing the THCAS-Like PCR product. As a result, a decision was made to replace the 366 SNP with the 1179 SNP on the THCAS gene (THCA1179), which was also expected to have the ability to differentiate marijuana samples from hemp seed samples, as supported by sequencing results from previous study (26). The updated differentiation flowchart based on the findings of this study is presented in Figure 1. Additionally, due to the close proximity of the 1064 SNP and the new 1179 SNP on the THCAS gene, the PCR primer was redesigned to simultaneously capture both SNP regions. The final PCR primer, SBE probes, and their optimized concentrations are provided in Table 2. An exemplary SNP profile demonstrating all possible alleles is shown in Figure 2. The interpretation of the genotype result will be explained in detail in the subsequence section along with Table 3. All 129 *C. sativa* samples tested successfully generated electropherograms with well-balanced peak heights while no peaks were observed in the one *Humulus lupulus* (hops) sample tested (Supplementary Table S3). Hops is a close relative of *C. sativa* (45) that was tested to validate the species specificity of the assay.

**Interpretation of** **SNaPshot™ genotyping assay**

Due to the complex relationship between the cannabis chemotype and genotype, the interpretation of genotyping results deviated from the traditional SNaPshotTM interpretation, where each allele combination typically receives an assigned genotype. In accordance with the differentiation flowchart (Figure 1), only five distinct genotypes were necessary for interpretation, with not all markers being employed in every genotype (Table 3, Supplementary Table S3). Data interpretation should start with the THCA1064 marker. If the genotype for this marker was either 'A' (genotype 1) or absent (genotypes 2), the sample could be classified as either a cannabis plant with ∆9-THC content <1% or a hemp seed. In such cases, the remaining markers held no significance. However, when the THCA 1064 marker indicated the presence of SNP G, a systematic evaluation of the other markers was necessary. In scenarios where the pseudo486 gene exhibited a solitary 'G,' the sample was typically identified as marijuana (genotype 3). The presence or absence of the CBDA586 peak could further distinguish between high CBD (present) and low CBD (absent) marijuana. However, an exception arose when the sample did not represent marijuana, specifically when the genotype combination of CBDA586 and THCA1179 was 'G' and 'A/T', respectfully. In such cases, the sample was categorized as a hemp seed sample (genotype 4). Lastly, in cases where either the pseudo486 marker or the THCA1064 exhibited heterozygosity (G/T or G/A), in conjunction with 'G' at the other marker, or if both markers showed heterozygosity, an examination of the peak height ratio (PHR) between the two 'G' SNPs was necessary (genotype 5). Genotype 5 with a PHR (1064G/486G) >1 was identified as marijuana, whereas a PHR <1 indicated a cannabinoid sample with THC content <1%. This SNaPshotTM assay was able to correctly interpret all 130 samples. Supplementary Figure S3 shows illustrative examples of each genotype and their corresponding interpretations.

The developed 4-plex SNaPshotTM assay successfully differentiated 48 hemp samples, encompassing 8 hemp types (CBD hemp, CBG hemp, Δ8-THC hemp, Δ10-THC hemp, THCO hemp, CBDV hemp, reference hemp, and hemp seed), from 81 marijuana samples, representing 7 marijuana types (DEA seized marijuana, Mexico seized marijuana flower, Mexico seized marijuana stem, Mexico seized marijuana seed, Chile marijuana, Chile medical marijuana, and reference marijuana). However, a limitation of the assay was its inability to distinguish hemp and marijuana based on the 0.3% legal threshold; instead, it could only differentiate marijuana samples with THC levels exceeding 1%. Consequently, 9 marijuana samples with THC levels below 1% could not be distinguished from hemp (Supplementary Table S3). Interestingly, the 1% THC threshold aligns with the criteria utilized in various seized drug chemistry analysis protocols, including those employed in crime labs by DEA and Texas crime laboratories (46, 47), for differentiating hemp from marijuana. Moreover, we noticed that excluding the CBG hemp and seed sample, all hemp has the pseodo486 T or G/T genotype (Supplementary Table S3) and genotype 5 heavily depends on the presence of THCAS-Like markers to facilitate the PHR approach for distinguishing specific hemp samples from marijuana. The expression pattern and function of the synthase pseudogene remain largely unexplored in the existing scientific literature (25). Our study presents a new perspective on the role of this pseudogene in relation to chemotype determination. This insight emphasizes the pivotal role played by pseudogenes in influencing cannabinoid content. This detailed understanding not only illuminates the complexities inherent in distinguishing hemp from marijuana but also underscores the need for precise genetic analyses to effectively categorize these diverse cannabis strains.

Overall, the SNaPshotTM assay testing boasts a well-rounded representation of various genotypes. However, genotype 4 was underrepresented, with just one sample falling into this category (Table 3). This low representation attributed to the limited availability of seed samples for analysis (Supplementary Table S1). Further investigations are needed to comprehensively assess the significance of the CBD586 and THCA1179 SNP in distinguishing hemp seeds from marijuana samples. Lastly, it is crucial to emphasize the significance of independently validating the analytical threshold in the laboratory when implementing the SNaPshotTM assay. Given the higher DNA input utilized in this study (2ng), the analytical threshold was intentionally set higher (200 RFUs) to prevent potential interference from trace amounts of SNP alleles, ensuring accurate interpretation of the results. Validation of the threshold is essential to maintain the assay's reliability and precision, particularly when dealing with samples containing varying DNA concentrations.

**TaqMan™ real-time PCR SNP genotyping assay design**

In this study, a highly specialized PCR-based TaqManTM real-time PCR SNP genotyping assay was designed to distinguish between hemp and marijuana. This assay was designed to detect two SNPs–the 1064 SNP situated on the THCAS gene and the 486 SNP on the THCAS-Like 5 gene. To achieve this, primer sets, THCAS-1064 and THCAS-Like 5, as detailed in Table 1 and Table 2, respectively, were selected to be utilized in the real-time PCR assay, effectively targeting the identified SNPs (Table 4). Adhering to the manufacturer's recommendation that the Tm of TaqMan® probes should be approximately 10°C higher than that of the primers (42), extended sequences were incorporated during the probe design. However, due to potential primer-probe interactions, the achieved Tm differences were around 6°C (Table 4). The TaqMan® QSY™ probes utilized in this assay feature a reporter dye at the 5´ end and a QSY™ quencher at the 3´ end, with four distinct reporter dyes – FAM™, VIC™, ABY™, and JUN™ – optimized for minimal spectral overlap, enabling simultaneous amplification of up to four targets in a single reaction. For the SNP at position 486 on the THCAS-Like 5 gene, with lower expression levels compared to the THCAS gene, FAM™ and VIC™ dyes were used for low to medium expressors, while ABY™ and JUN™ dyes were employed for medium to high expressors targeting the 1064 SNP on the THCAS gene. The final PCR primer, TaqMan® probes, and their optimized concentrations are provided in Table 4.

The assay was applied to the same *C. sativa* (n=128) and hop (n=1) samples as the SNaPshotTM assay with one medical marijuana sample from Chile was excluded due to its complete consumption during the previous run (Supplementary Table S4). Five samples with known genotypes were designated as positive controls. The combined allelic discrimination plots for all sample are showed in Figure 3. The interpretation of the genotype result will be explained in detail in the subsequence section along with Table 5.

**Interpretation of TaqManTM real-time PCR SNP genotyping assay**

Similar to the SNaPshotTM assay, initial data interpretation should start with the THCA1064 marker. Samples exhibiting genotype cluster of either A/A (genotype 1) or G/A (genotypes 2) at this marker are indicative of cannabis plants with ∆9-THC content <1%. Conversely, samples clustering around the G/G alleles for the THCA1064 marker suggest marijuana samples. However, if sample cluster at the undetermined area for the THCA1064 marker, it is necessary to evaluate of the other marker, pseudo486. However, in cases where sample genotype was undetermined for the THCA1064 marker, further evaluation involving the pseudo486 marker becomes necessary (Table 5). When samples cluster at T/T, T/G, or within the undetermined area for the pseudo486 marker, they were identified as cannabis plants with ∆9-THC content <1% (genotypes 4, 5, and 6). Conversely, the presence of the G/G allele at the pseudo486 marker indicates the presence of marijuana (Table 5). As anticipated, the absence of the CBDA586 and THCA1179 markers in the differentiation flowchart for the Taqman assay results in an inability to distinguish hemp seeds from marijuana samples, which showing the same genotypes as genotypes 3 and 7. Furthermore, non-cannabis plant material will exhibit non-amplification of both markers, leading to classification as genotype 6, indicative of cannabis plants with ∆9-THC content <1% (Table 5).

The developed multiplex TaqManTM real-time PCR SNP genotyping assay successfully differentiated all commercial hemp flowers and reference hemp samples from marijuana samples. Similar to the SNaPshot™ assay, the assay was unable to differentiate marijuana samples with THC levels between 0.3% and 1% from hemp samples. Consequently, 12 marijuana samples with THC levels below 1% could not be distinguished from hemp (Supplementary Table S4). This indicates a limitation in the assay's sensitivity to THC concentration within this range, which is critical for regulatory purposes. Additionally, we observed that nine marijuana samples from Chile and Mexico were also classified as Cannabis with THC < 1% according to our assay. Unfortunately, we could not obtain the chemical data for these samples to validate our findings. Furthermore, one marijuana sample (DEA 100C) with THC > 1% was misclassified as Cannabis with THC < 1%, indicating a potential issue with assay accuracy at specific THC thresholds.

Moreover, the assay exhibited limitations in distinguishing hemp seeds from marijuana samples due to the absent of the THCAS1179 and CBDAS586 markers. The assay also showed limitations in differentiating non-cannabis samples from hemp. The non-cannabis sample yielded undetermined results with both markers and clustered with hemp samples showing similar results, suggesting that the assay may not effectively distinguish non-cannabis from hemp. Despite these limitations, the assay remains valuable for its primary application, particularly given the legality of hemp in the USA and the relative rarity of hemp seed samples compared to other types. The ability to accurately identify commercial hemp and reference hemp flowers from marijuana is significant for regulatory compliance and industry practices, even though further refinement is needed to improve differentiation at lower THC levels and among non-cannabis samples.

The advantages and limitations of these assays are summarized in Table 6. Despite their respective limitations, both the SNaPshot™ and TaqMan™ assays offer rapid and cost-effective genotyping solutions for distinguishing hemp from marijuana by targeting specific SNP markers. The SNaPshot™ assay requires only standard forensic laboratory equipment, such as a PCR and a CE instrument, with an estimated run time of 3.5 to 4 hours for processing 96 samples. Conversely, the TaqMan™ real-time PCR assay requires only an RT-PCR instrument and has a total run time of 1.5 hours for 96 samples. Both assays provide higher throughput compared to traditional chemical methods.

**Conclusion**

The utilization of genetic tools to differentiate between hemp and marijuana serves as a valuable alternative to conventional chemistry methods, especially when dealing with samples unsuitable for chemical analysis. Moreover, it addresses issues related to variability caused by sample storage and aging. A recent study involved the development of a comprehensive differentiation flowchart utilizing four markers through NGS (26). In this study, we applied the previously proposed markers to develop two methodologies applicable in most crime laboratories for the differentiation of hemp and marijuana. Although each with its unique strengths and limitations, the two assays stand as valuable tools for crime labs, addressing the practical needs of rapid and accurate cannabis categorization. An updated identification flowchart for distinguishing between hemp and marijuana was proposed, providing a comprehensive framework for classification. Utilizing this updated flowchart, the developed SNaPshotTM and TaqMan™ assays demonstrated discrimination abilities across most hemp categories and marijuana samples, except for samples with less than 1% THC. This study advances our understanding of hemp-marijuana divergence and the potential role of the THCAS pseudogene related to the cannabis chemotype. This work provides step toward enhancing the efficiency and reliability of forensic cannabis analysis, emphasizing the indispensable role of genetic insights in the evolving landscape of cannabis regulation and enforcement.

**Acknowledgements**

This project was partially funded by Award # 1739805 through the Center for Advanced Research in Forensic Science (CARFS), an NSF-funded IUCRC at Florida International University. The opinions, findings, conclusions, or recommendations expressed in this article are those of the authors. The authors would like to thank the National Institute on Drug Abuse (NIDA), National Institute of Standards and Technology (NIST) for providing marijuana specimens for this project. Additionally, the authors would like to thank the University of Mississippi for their kind donation of *Cannabis* DNA extracts and Dr. Ann-Elodie Black for providing cannabinoid chemistry data of those samples.

**References**

1. United State Congress. H.R.2-Agricultural Improvement Act of 2018. https://www.govinfo.gov/content/pkg/PLAW-115publ334/pdf/PLAW-115publ334.pdf. Accessed 25 May 2024.

2. Johnson R. Defining Hemp: A Fact Sheet. Washington, D.C.: Congressional Research Service; 2019 Mar. Report No.: R44742.

3. Brenneisen R, elSohly MA. Chromatographic and spectroscopic profiles of Cannabis of different origins: Part I. J Forensic Sci. 1988 Nov;33(6):1385-404.

4. Shibuya EK, Souza Sarkis JE, Neto ON, Moreira MZ, Victoria RL. Sourcing Brazilian marijuana by applying IRMS analysis to seized samples. Forensic Sci Int. 2006 Jun 27;160(1):35-43.

5. Dufresnes C, Jan C, Bienert F, Goudet J, Fumagalli L. Broad-Scale Genetic Diversity of Cannabis for Forensic Applications. PLoS One. 2017;12(1):e0170522.

6. Kojoma M, Seki H, Yoshida S, Muranaka T. DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in "drug-type" and "fiber-type" Cannabis sativa L. Forensic Sci Int. 2006 Jun 2;159(2-3):132-40.

7. Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, Butler L, et al. The Genetic Structure of Marijuana and Hemp. PLoS One. 2015;10(8):e0133292.

8. Gilmore S, Peakall R, Robertson J. Organelle DNA haplotypes reflect crop-use characteristics and geographic origins of Cannabis sativa. Forensic Sci Int. 2007 Oct 25;172(2-3):179-90.

9. Houston R, Birck M, Hughes-Stamm S, Gangitano D. Evaluation of a 13-loci STR multiplex system for Cannabis sativa genetic identification. Int J Legal Med. 2016 May;130(3):635-47.

10. Cascini F, Farcomeni A, Migliorini D, Baldassarri L, Boschi I, Martello S, et al. Highly Predictive Genetic Markers Distinguish Drug-Type from Fiber-Type Cannabis sativa L. Plants (Basel, Switzerland). 2019;8(11):496.

11. Solano J, Anabalón L, Figueroa A, Gangitano D. ITS barcoding using high resolution melting analysis of Cannabis sativa drug seizures in Chile: A forensic application. Forensic Sci Int. 2020 Nov;316:110550.

12. Borin M, Palumbo F, Vannozzi A, Scariolo F, Sacilotto GB, Gazzola M, et al. Developing and Testing Molecular Markers in Cannabis sativa (Hemp) for Their Use in Variety and Dioecy Assessments. Plants (Basel). 2021 Oct 14;10(10).

13. Rotherham D, Harbison SA. Differentiation of drug and non-drug Cannabis using a single nucleotide polymorphism (SNP) assay. Forensic Sci Int. 2011 Apr 15;207(1-3):193-7.

14. Yamamuro T, Segawa H, Kuwayama K, Tsujikawa K, Kanamori T, Iwata YT. Rapid identification of drug-type and fiber-type cannabis by allele specific duplex PCR. Forensic Sci Int. 2021 Jan;318:110634.

15. Tahir MN, Shahbazi F, Rondeau-Gagné S, Trant JF. The biosynthesis of the cannabinoids. Journal of cannabis research. 2021;3(1):7-.

16. Liu Y, Zhu P, Cai S, Haughn G, Page JE. Three novel transcription factors involved in cannabinoid biosynthesis in Cannabis sativa L. Plant Mol Biol. 2021 May;106(1-2):49-65.

17. García-Valverde MT, Sánchez-Carnerero Callado C, Díaz-Liñán MC, Sánchez de Medina V, Hidalgo-García J, Nadal X, et al. Effect of temperature in the degradation of cannabinoids: From a brief residence in the gas chromatography inlet port to a longer period in thermal treatments. Front Chem. 2022;10:1038729.

18. Rotherham D, Harbison SA. Differentiation of drug and non-drug Cannabis using a single nucleotide polymorphism (SNP) assay. Forensic Science International. [Article]. 2011;207(1-3):193-7.

19. Roman MG, Cheng YC, Kerrigan S, Houston R. Evaluation of tetrahydrocannabinolic acid (THCA) synthase polymorphisms for distinguishing between marijuana and hemp. J Forensic Sci. 2022 Jul;67(4):1370-81.

20. Laverty KU, Stout JM, Sullivan MJ, Shah H, Gill N, Holbrook L, et al. A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. Genome research. 2019;29(1):146-56.

21. Fulvio F, Paris R, Montanari M, Citti C, Cilento V, Bassolino L, et al. Analysis of Sequence Variability and Transcriptional Profile of Cannabinoid synthase Genes in Cannabis sativa L. Chemotypes with a Focus on Cannabichromenic acid synthase. Plants (Basel). 2021 Sep 8;10(9).

22. Grassa CJ, Weiblen GD, Wenger JP, Dabney C, Poplawski SG, Timothy Motley S, et al. A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. New Phytol. 2021 May;230(4):1665-79.

23. Weiblen GD, Wenger JP, Craft KJ, ElSohly MA, Mehmedic Z, Treiber EL, et al. Gene duplication and divergence affecting drug content in Cannabis sativa. The New phytologist. 2015;208(4):1241-50.

24. Onofri C, de Meijer EPM, Mandolino G. Sequence heterogeneity of cannabidiolic- and tetrahydrocannabinolic acid-synthase in Cannabis sativa L. and its relationship with chemical phenotype. Phytochemistry. 2015 Aug;116:57-68.

25. Hurgobin B, Tamiru-Oli M, Welling MT, Doblin MS, Bacic A, Whelan J, et al. Recent advances in Cannabis sativa genomics research. New Phytol. 2021 Apr;230(1):73-89.

26. Cheng YC, Houston R. The development of a next-generation sequencing panel targeting cannabinoid synthase genes to distinguish between marijuana and hemp. Electrophoresis. 2024 May;45(9-10):948-57.

27. Nwawuba Stanley U, Mohammed Khadija A, Bukola AT, Omusi Precious I, Ayevbuomwan Davidson E. Forensic DNA Profiling: Autosomal Short Tandem Repeat as a Prominent Marker in Crime Investigation. Malays J Med Sci. 2020 Jul;27(4):22-35.

28. ABI PRISM® SNaPshot™Multiplex Kit Protocol 4323357b. Thermo Fisher Scientific, South San Francisco, CA; 2010. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms\_041203.pdf. Accessed 25 May 2024.

29. Fondevila M, Børsting C, Phillips C, de la Puente M, Consortium EN, Carracedo A, et al. Forensic SNP genotyping with SNaPshot: Technical considerations for the development and optimization of multiplexed SNP assays. Forensic Sci Rev. 2017 Jan;29(1):57-76.

30. Jawhari M, Abrahamian P, Sater AA, Sobh H, Tawidian P, Abou-Jawdah Y. Specific PCR and real-time PCR assays for detection and quantitation of 'Candidatus Phytoplasma phoenicium'. Mol Cell Probes. 2015 Feb;29(1):63-70.

31. Birdsell DN, Vogler AJ, Buchhagen J, Clare A, Kaufman E, Naumann A, et al. TaqMan real-time PCR assays for single-nucleotide polymorphisms which identify Francisella tularensis and its subspecies and subpopulations. PLoS One. 2014;9(9):e107964.

32. Haddrill PR. Developments in forensic DNA analysis. Emerg Top Life Sci. 2021 Sep 24;5(3):381-93.

33. Deepak S, Kottapalli K, Rakwal R, Oros G, Rangappa K, Iwahashi H, et al. Real-Time PCR: Revolutionizing Detection and Expression Analysis of Genes. Curr Genomics. 2007 Jun;8(4):234-51.

34. Schleinitz D, DiStefano JK, Kovacs P. Targeted SNP Genotyping Using the TaqMan® Assay. Totowa, NJ: Humana Press; 2011;77-87.

35. DNeasy® plant handbook. QIAGEN, Hilden, Germany; 2018. https://www.qiagen.com/us/resources/download.aspx?id=f6455f80-dc4f-4ff2-b2de-ae7a3e6c91e0&lang=en. Accessed 25 May 2024.

36. Qubit® dsDNA HS Assay Kits Protocol. Thermo Fisher  Scientific, South San Francisco, CA; 2015. https://tools.thermofisher.com/content/sfs/manuals/Qubit\_dsDNA\_HS\_Assay\_UG.pdf. Accessed 25 May 2024.

37. Houston R, Birck M, LaRue B, Hughes-Stamm S, Gangitano D. Nuclear, chloroplast, and mitochondrial data of a US cannabis DNA database. Int J Legal Med. 2018 May;132(3):713-25.

38. Multiple Primer Analyzer. Thermo Fisher Scientific; 2006. <https://www.thermofisher.com/us/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html>. Accessed 25 May 2024.

39. Type-it Microsatellite PCR Handbook. QIAGEN, Hilden, Germany; 2009. https://www.qiagen.com/us/resources/resourcedetail?id=d6135896-0466-4d9d-aeef-1c4f23f8964e&lang=en. Accessed 25 May 2024.

40. BigDye™ Terminator v3.1 Cycle Sequencing Kit User Guide. Thermo Fisher  Scientific, South San Francisco, CA; 2016. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/4337035\_BDTv31CycSqKt\_RUO\_UG.pdf. Accessed 25 May 2024.

41. TaqMan QSY probes. South San Francisco, CA: Thermo Fisher Scientific; 2021. <https://assets.thermofisher.com/TFS-Assets/GSD/Flyers/taqman-qsy-probes-qpcr-multiplexing-flyer.pdf>. Accessed 25 May 2024.

42. TaqMan™ Assay Multiplex PCR Optimization-APPLICATION GUIDE. South San Francisco, CA: Thermo Fisher Scientific; 2022. <https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2Ftaqman_optimization_man.pdf>. Accessed 25 May 2024.

43. TaqPath™  ProAmp™ Master Mixes-USER GUIDE. South San Francisco, CA: Thermo Fisher Scientific; 2022. <https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2FMAN0015758_TaqPathProAmpMMix_UG.pdf>. Accessed 25 May 2024.

44. McKernan K, Helbert Y, Kane L, Ebling H, Zhang L, Liu B, et al. Sequence and annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes, 2020.

45. McPartland JM. Cannabis Systematics at the Levels of Family, Genus, and Species. Cannabis Cannabinoid Res. 2018;3(1):203-12.

46. Drug Enforcement Administration. Analysis of Drug Manual. Office of Forensic Science; 2019. https://www.dea.gov/sites/default/files/2019 10/Forensics/ADM%20R4%202019\_Public%20Posting\_Final2.pdf. Accessed 25 May 2024.

47. DPS T. DPS Crime Laboratory-FAQ for Identification of Marihuana Post Hemp Legalization. Seized Drugs Section; 2020. https://txdpslabs.qualtraxcloud.com/showdocument.aspx?ID=89573. Accessed 25 May 2024.

**Tabl****es**

**Figure Captions**

**Figure 1.** Updated differentiation flow chart. The red boxes were markers modified in present study

**Figure 2.** Four-plex SNaPshotTM assay with all possible alleles

**Figure 3.** Combined allelic discrimination plots for (a) THCA1064 and (b) pseudo486 markers. The X and Y axes represent ΔRn values