International Journal of Agricultural Technology 2023 Vol. 19(4):1891-1906
Available online http://www.ijat-aatsea.com
ISSN 2630-0192 (Online)

Evaluation of SNP markers linked to the *THCA* and *CBDA* synthase gene for varietal improvement of *Cannabis sativa* L.

Tawinchure, T.[1], Toterakun, T.[2], Rumjuankiat, K.[1] and Urairong, H.[1*]

[1]College of Agricultural Innovation and Food Technology, Rangsit University, Pathum Thani, Thailand; [2]Plants R&D Manager Mae Fah Luang Foundation, Doi Tung Development Project, Chiang Rai, Thailand.

Tawinchure, T., Toterakun, T., Rumjuankiat, K. and Urairong, H. (2023). Evaluation of SNP markers linked to the *THCA* and *CBDA* synthase gene for varietal improvement of *Cannabis sativa* L. International Journal of Agricultural Technology 19(4):1891-1906.

Abstract This study compared the nucleotide variation of the *THCAS* gene between the high THC (Type I) and high CBD (Type III) of cannabis plants. The results showed that 12 of 63 SNP positions were selected as having the potential to distinguish between these two types. Validation of the 12 SNPs was performed in F1 progeny derived from the cross between Type III (DNM) and Type I (CHP) plants using a pyrosequencing technique and *in silico* analysis. The results revealed that the SNP positions 869 (T/C) and 881 (T/G) could be used as effective markers to predict the cannabinoid type of cannabis plant correctly but not the quantity. Finally, the tetra-primer ARMS-PCR was designed to be used as the Marker-Aided Selection in the cannabis breeding program.

Keywords: Cannabis, SNPs, THC, CBD, Tetra-primer ARMS-PCR

Introduction

Cannabis sativa L., popularly known as marijuana (drug type) and hemp (fiber type), is commonly called "cannabis." Cannabis is a dioecious plant with separate male and female flowers on individual plants, and cannabis is an open-pollinated plant (Salentijn *et al.*, 2015). This plant has been cultivated worldwide for thousands of years for the therapeutic and psychoactive properties of cannabinoids; cannabinoids are the major organic compound in cannabis. Nowadays, Thailand has become the first Asian nation to legalize medical cannabis. As a result, many legal farms and companies in Thailand started widely growing cannabis and developing products such as hemp seed oil, CBD oil, THC oil, CBD drink and beverage, and cosmetics. Nevertheless, Thai domestic cannabis has not been bred for a long time because of the illegalization in the past, causing unimprovement of the genotype and phenotype of Thai domestic cannabis by appropriate varietal improvement programs until today.

Cannabinoids are accumulated in the trichrome of cannabis inflorescences, containing approximately 150 compounds that have been

* **Corresponding Author:** Urairong, H.; **Email:** fongptt@yahoo.com

identified. Two significant cannabinoids in medical and recreational uses are tetrahydrocannabinol (Δ 9-THC) and cannabidiol (CBD). THC is a psychoactive compound with antiemetic and pain-killer effects. While CBD is a non-psychoactive compound with interesting medical properties such as antiepileptic, anti-inflammatory, and analgesic activities (Ramirez *et al.*, 2019). THC and CBD have the same precursor, which is cannabigerolic acid (CBGA) which is converted into tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA) by the metabolic of THCA synthase and CBDA synthase, respectively. At the final biosynthetic pathway, THCA and CBDA were decarboxylated and converted into pharmacologically active forms of THC and CBD by heat (Sirikantaramas and Taura, 2017). Cannabis plants produce various cannabinoid content in different cannabis lines. All cannabis lines have been classified into five broad chemotypes based on proportional ratios of THCA, CBDA, and CBGA. "Type I" cannabis plants contain predominantly THCA and low CBDA content, "Type II" plants with an approximately equal ratio of THCA and CBDA (intermediate chemotype), "Type III" plants, with high CBDA level and relatively low THCA level, "Type IV" plants contain CBGA which is the precursor of THCA and CBDA, and "Type V" plants are the cannabinoid-free type that undetectable amounts of CBGA, THCA, and CBDA (Garfinkel *et al.*, 2021). The Analysis of cannabinoid concentration plays a key role in many fields of study and industries, especially in cannabis breeding programs to select the appropriate offspring with desired cannabinoid content. However, analyzing the cannabinoid content of every offspring plant by gas chromatograph (GC), gas chromatography-flame ionization detector (GC-FID), or high-performance liquid chromatography (HPLC) might be extravagant methods.

To breed new cannabis lines requires a specific breeding approach that comprises 3 phases (1) searching for the natural variation in the material and creating a base population, (2) generating varietal parents through selection and improving the population through recurrent selection steps to create a breeding population and (3) development and testing experimental cultivars (Posselt, 2010). Moreover, the important method for cannabis breeding has four standard processes, which are mass selection, cross-breeding, inbreeding, and hybrid breeding (Ranalli, 2004). Still, traditional breeding takes considerable time and the cost of HPLC analysis for offspring selection is expensive. At the same time, molecular markers have been developed to assist breeders in making decisions for progeny selection, trait characterizations, and genetic manipulation. These useful markers were developed to minimize extensive phenotyping, which could be expensive and time-consuming (Singh *et al.,* 2021). Single nucleotide polymorphism (SNP) is a different single base in DNA among individuals. Various types of SNPs could change the function or the regulation and expression of a protein (Brooks, 2003). Kojoma *et al.* (2006) examined the SNPs that could

be markers to distinguish between Type I and Type III cannabis for forensic and breeding applications. After that, Cascini *et al*. (2019) studied more about the SNP positions in *THCAS* and *CBDAS* genes. They explained the biochemical mechanisms that regulate the cannabis chemical profile using an *in silico* analysis to develop an effective marker for forensic investigations. Moreover, the applications of SNPs marker in cannabis breeding could assist the breeders in having early selection at the seedling stage. The marker could be applied to select the desired chemotype of the offspring by using only the DNA of young leaves. Also, using the marker could save much cost and considerable time for the cannabis varietal improvement program. Many PCR-based techniques could detect the SNP, such as real-time PCR, allele-specific detection by amplification refractory mutation system (ARMS) analysis, and genotyping by melting curve analysis (Matsuda, 2017). Genotyping by pyrosequencing is the alternative de novo DNA sequencing technology that could detect the SNP, also the insertion and deletion of nucleotide (InDel). This technique is based on the sequencing by synthesis principle, using an enzyme cascade system comprising DNA polymerase, ATP sulfurylase, luciferase, and apyrase to accurately detect nucleic acid sequences (Fakruddin and Chowdhury, 2012). Furthermore, the tetra-primer ARMS-PCR was optimized as a rapid and economical technique to detect the SNPs. The tetra-primer ARMS-PCR uses two sets of primers. The first set was the outer primers, which were used to amplify the target region of DNA sequence with the same function as conventional PCR primers. The second set was two inner primers, which were used to detect the allelic variation. The product of tetra-primer ARMS-PCR could be analyzed using agarose gel electrophoresis (Sundru *et al.,* 2015)

This research aimed to detect the SNP positions in *THCAS* or *CBDAS* genes to predict the chemotype of the cannabis plant and to utilize the marker aided for varietal Improvement of *Cannabis sativa* L.

Materials and methods

Plant materials

This experiment uses eight Type I Thai domestic lines, one Type I commercial line, and four Type III commercial lines. All commercial lines have a certificate of analysis (COA) to verify the cannabinoid content in their female inflorescences. Moreover, the cross between DNM x CHP was made at Plants R&D Manager Mae Fah Luang Foundation, Doi Tung Development Project, Thailand. Therefore, this research used these samples for SNPs analysis, as shown in Table 1.

Table 1. List of cannabis lines and chemotype used in this study

Code/line	Chemotype	Source	Strain
Wild Thai (WT)	I	Thai domestic	Sativa
Lopburi-1 (LOP-1)	I	Thai domestic	Sativa
Lopburi-2 (LOP-2)	I	Thai domestic	Sativa
Suphan (SP)	I	Thai domestic	Sativa
Phrik Thai (PT)	I	Thai domestic	Sativa
Chang Phuak[2] (CHP)	I	Thai domestic	Sativa
Kroeng Krawia (KRKW)	I	Thai domestic	Sativa
KD	I	Thai domestic	Sativa
King Kong (KK)	I	Commercial line	Sativa 20% x Indica 80%
BRR	III	Commercial line	Sativa x Indica
Dinamed[1] (DNM)	III	Commercial line	Sativa 60% x Indica 40%
Crystal Med (CRST)	III	Commercial line	Sativa 50% x Indica 50%
Charlotte's Angel (CHLA)	III	Commercial line	Sativa 60% x Indica 40%
DCP22 (10 plants)	Unknown	Hybridized in this study	Dinamed x Chang Phuak

[1/] Female parent of DCP22, [2/]Male parent of DCP22

Detection of SNP positions in THCAS and CBDAS genes

All full-length *THCAS* gene nucleotide sequences were retrieved from the GenBank of the National Center for Biotechnology Information (NCBI: https://www.ncbi.nlm.nih.gov). The five *THCAS* gene sequences derived from high THC plants (Type I), accession number MG996415, MG996416, MW504063, MG996401, MW504064, were compared with another five sequences derived from high CBD plants (Type III), accession number AB212830, AB212833, MG996403, MG996406, and MG996408. The multiple alignments of these ten sequences were accomplished to select the SNP positions which could distinguish between Type I and Type III cannabis using MEGA X software (https://www.megasoftware.net). Simultaneously, the same criteria were done to the *CBDAS* gene, five full-length nucleotide sequences of the *CBDAS* gene were retrieved from Type III plants, accession numbers KJ469374, KP970861, KP970865, MG996431, and MG996437 were compared with another five sequences of Type I plants, accession number MG996420, MG996423, MG996425, MG996426, MG996427.

DNA extraction

Briefly, fresh leaves of samples were collected from vegetative stage cannabis for DNA extraction. Total genomic DNA was extracted from 100

mg of fresh leaves using the modified cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). The concentration and purity of DNA were determined using the ratio of absorbance at 260/280 nm by UV/Vis Spectrophotometer (Libra S22; Biochrom, UK). The final concentration of DNA was adjusted to 50 ng/µl with 1X TE buffer and stored at -20 ℃ for further study.

SNPs validation by pyrosequencing

The pyrosequencing technique was used to validate the SNP positions that we selected from the multiple alignment data. All the aligned sequences are derived from different origins of cannabis. However, the prediction potential of those SNP positions needed to be confirmed in Thai domestic cannabis lines. The pyrosequencing analysis was divided into 3 steps, i) Primer design: PyroMark® Assay Design software (QIAGEN, Ger) was used for designing biotinylated PCR primers and sequencing primers, the primers, chemicals and enzymes related to pyrosequencing reaction were purchased from QIAGEN ii) The PCR amplification was performed in a total volume of 25 µl, the reaction of each sample containing 2X PyroMark PCR Master Mix, 10X CoralLoad Concentrate, 0.2 µM of forward and reverse primers, RNase-free water, and 2 ng of DNA template. The PCR condition was preheating at 95 ℃ for 15 minutes, 45 cycles of denaturation at 94 ℃ for 30 seconds, annealing at 60 ℃ for 30 seconds, extension at 72 ℃ for 30 seconds, and final extension at 72 ℃ for 10 minutes using T100 Thermal Cycler (Bio-Rad, USA). The PCR products were electrophoresed on 1.5% agarose gel then stained with 5% SafeView™ Classic (abm®, CA), and examined under blue light LED (470 nm) using B-BOX™ Blue Light LED Epi-illuminator (SMOBIO®, TW). iii) PCR products were analyzed by PyroMark® Q48 Autoprep machine (QIAGEN, Ger) to detect and analyze the SNPs of each gene. The pyrosequencing reaction contained 10 µl biotinylated PCR product, and 3 µl PyroMark Q48 magnetic beads were loaded into PyroMark Q48 Disc before placing into the machine. PyroMark Q48 Advanced Reagents contain PyroMark binding buffer, enzyme and substrate mixture, and sequencing primer (final concentration 4 µM).

Prediction of protein function and effect of single amino acid substitution on protein

The nucleotide sequences of *THCAS* and *CBDAS* genes that contained selected SNP positions were translated into amino acid sequences using MEGA X software. Consequently, the amino acid sequences were subjected to multiple alignments to verify the SNPs impact on the protein's function. The effect of single amino acid substitution (SAAS) on protein function was analyzed by *in silico* software named Plant Protein Variation

Effect Detector (PPVED), PPVED can be accessed under a CC-BY (4.0) license via http://www.ppved.org.cn.

THC and CBD content analysis by HPLC

All mature inflorescence of cannabis samples were decarboxylated and ground before being filtered using a 60 (0.0098 inches) test sieve (Cole-Parmer®, USA) before the analysis. The mobile phase A of HPLC was ammonium formate:acetonitrile that was prepared by dissolving 3.153g of ammonium formate in 10% v/v acetonitrile to 1,000 ml, pH 3.75, whereas mobile phase B was 90% v/v acetonitrile. The gradient elution HPLC was performed by using the XBridge C18 column (3.5 µm, 4.6 x 150 nm) with a flow rate of 1.0 mL/min for 40 minutes to detect and analyze the concentration of THC and CBD.

Progeny analysis to verify the highly predictive SNP positions

A cross between DNM was made, homozygous CBD dominant plant (DNM) as a female parent with male homozygous THC dominant plant (CHP) (Table 3.). The leaves of the F1 seedling stage plants were collected for DNA extraction. The SNP genotyping of *THCAS* and *CBDAS* genes was determined by pyrosequencing. The mature inflorescence of each F1 plant was analyzed for cannabinoid content by HPLC.

Tetra-primer ARMS-PCR

The tetra-primer ARMS-PCR technique was used to detect the SNP, which has high discrimination potential between Type I and Type III cannabis. Primers were designed by Primer1 (primer1.soton.ac.uk/primer1.html). The amplification was performed in a total volume of 20 µl. The PCR reaction of each sample contained 2X DreamTaq Green PCR Master Mix (Thermo Scientific, USA), DNA template, RNase-free water, and four primers which are forward outer (FO), reverse outer (RO), forward inner (FI), and reverse inner (RI) primer. The tetra-primer ARMS-PCR products were examined by electrophoresis using 3% agarose gel, staining with 5% SafeView™, and examined under blue light LED (470 nm) using B-BOX™ Blue Light LED Epi-illuminator (SMOBIO®, TW).

Results

Detection of SNP position in THCAS and CBDAS genes

The nucleotide multiple alignment sequence of the *THCAS* gene, a comparison between Type I and Type III cannabis, was found 63 SNP positions that could distinguish between Type I and Type III plants. Moreover, more sequences were added to multiple alignments to confirm the effective SNPs. Hence, 12 of 63 SNP positions were selected as having high discrimination potential, as 612, 678, 699, 744, 749, 763, 862, 864, 869, 881, 885, and 887. These positions were used to design sequencing primers for pyrosequencing. At the same time, 70 SNP positions that could distinguish between Type I and Type III plants were found in the *CBDAS* gene (3 bp and 4 bp InDel included). Thus, the same effective SNPs confirmation criteria were done to the *CBDAS* gene, 5 SNPs and 3 bp InDel in the *CBDAS* gene at positions 583, 584, 588, 613, 637, and 758 (3 bp InDel) were selected by the same criteria as the selection of *THCAS* gene.

SNPs validation by pyrosequencing

The biotinylated PCR primers (FP-THCA2, PRB-THCA2) were designed and used to amplify the specific region between nucleotide positions 549-931 in the *THCAS* gene (Table 2.) In this target region, four sequencing primers were designed to detect the percentage of single nucleotide substitution of 12 SNP positions (Table 2.). For the *CBDAS* gene, the PCR primers of two regions (FP-CBDA1, RPB-CBDA1 and FP-CBDA2, RPB-CBDA2) were designed to amplify between nucleotide positions 451-735 and 690the -960 of *CBDAS* gene, and the two sequencing primers were designed to detect the percentage of single nucleotide substitution of 5 SNP positions and 3 bp InDel (Table 2).

In the comparison of 12 SNP positions between high THC plants (Type I) and high CBD plants (Type III) in the *THCAS* gene (Table 3.), the results showed that positions 869 and 881 have a highly distinguished potential between Type I and Type III cannabis, other positions showed alleles in some cannabis lines randomly. Moreover, the advantage of the pyrosequencing technique was not only to detect the SNP but also could analyze the percentage of single nucleotide substitution as shown in SNP position 869 results. For example, WT, LOP-1, LOP-2, SP, PT and CHP showed both T alleles (homozygous) at 82%, 86%, 86%, 87%, 86%, and 94%, respectively. Furthermore, DNM showed homozygous C alleles at 83%. Nevertheless, BRR, CRST, CHLA showed 2 alleles of T and C at different percentages of each line. Finally, we conclude that T allele of SNP position 869 in *THCAS* gene could predict the chemotype of cannabis as Type I plant, while the C allele represents the Type III plant at position 869. For the *CBDAS* gene, the comparison of 5 SNP positions and 3 bp InDel position (Table 4.) the results indicated that SNP position 584 could distinguish between Type I and Type III plants, at position 584 all Thai domestic lines showed homozygous T alleles, while Type III plants showed

homozygous G alleles except for CRST and CHLA which are Type III plants but showed both T alleles. T allele represent Type I plant while G allele represent Type III plant. Moreover, at position 758 (3 bp InDel), all Type I plants showed the 3 bp insertion of CAA bases except for LOP-2 which showed the 3 bp deletion (- - -) while Type III plants showed 3 bp insertion except CRST and CHLA that showed 3 bp deletion. Subsequently, the prediction of protein translation was analyzed and the results indicated that the SNPs of *THCAS* gene at positions 749 (C/T), 763 (T/G), 862 (G/A), 869 (T/C), 881 (T/G), and 887 (A/G) were non-synonymous SNPs, which caused single amino acid substitution in amino acid sequence at positions 250 (Alanine (A) → Valine (V)), 255 (Serine (S) → Alanine (A)), 288 (Valine (V) → Methionine (M)), 290 (Methionine (M) → Threonine (T)), 294 (Isoleucine (I) →Alanine (A)) and 296 (Lysine (K) → Arginine (R)), respectively. For *CBDAS* gene, the results found that the SNPs at positions 583 (A/C), 584 (G/T), 613 (A/G), and 637 (C/G) cause the codon changing at positions 195 (Arginine (R) → Isoleucine (I)), 195 (Arginine (R) → Leucine (L)), 205 (Isoleucine (I) → Valine (V)), 213 (Histidine (H) → Aspartic acid (D)) of amino acid sequence, respectively.

Prediction of protein and effect of single amino acid substitution on protein function

The effect of SAAS on each SNPs position was analyzed using the PPVED program. The results of *THCAS* gene (Table 3.) demonstrated the synonymous SNPs at positions 612, 678, 699, 744, and 885, respectively, which do not cause the changes in the amino acid at these positions. For this reason, those synonymous positions were excluded from the focus group of this research. Furthermore, the results indicated that positions 749 (C/T), 763 (T/G), 862 (G/A), and 887 (A/G) were predicted as the neutral class, which means the protein function at these positions were not changed, and remain the same function, while SNPs positions 869 (T/C) and 881 (T/G) were predicted as a functional class, which means the codon changing causes the changes of protein function at these two positions. According to the prediction results, we consider focusing on SNPs that were functional class. Thus, the SNP position 869 was selected for tetra-primer ARMS-PCR design because of the classification as the functional class and the detection of heterozygous in other lines (KRKW, KK, BRR, CRST, and CHLA). The detection of heterozygous plays a key role in marker-aided selection for plant breeding program. On the other hand, SNPs of the *CBDAS* gene (Table 4.). We found synonymous SNP at position 588, and non-synonymous SNPs at positions 583 (A/C), 584 (G/T), 613 (A/G), and 637 (C/G). The prediction results indicated that positions 583 and 584 were predicted as a functional class, but positions 583 and 584 could not completely distinguish Type I from Type III cannabis as the results of

CRST and CHLA. Moreover, SNP positions 613 (A/G) and 637 (C/G) were classified as the neutral class hence all the SNPs of the *CBDAS* gene were excluded from the focus group of this research.

Table 2. Oligonucleotide sequence of PCR and sequencing primers for the pyrosequencing of *THCAS* and *CBDAS* genes

Target gene	Name	Primer Sequence (5'-3')	Target region/position
THCAS	PCR primer		
	FP-THCA2	TGGCGTAGGTGGACACTTTA	549-931
	RPB-THCA2	Biotin-TCTTATTCTTCCCATGATTATCTG	
	Sequencing primer		
	SP-THCA2-1	CCTTGCGGCTGATAA	612
	SP-THCA2-2	AGTTCTAGATCGAAAATCC	678, 699
	SP-THCA2-3	AGCATGGAAAATCAAAC	744, 749, 763
	SP-THCA2-4	AATATTGCTTACAAGTATGA	862, 864, 869, 881, 885, 887
CBDAS	PCR primer		
	FP-CBDA1	GGAGCTACCCTTGGAGAAGTTTA	451-735
	RPB-CBDA1	Biotin-GGGACAGCAACCAGTCTAATT	
	FP-CBDA2	ACGTGGTGGTGGAGCAGAAA	690-960
	RPB-CBDA2	Biotin-CACTCCACCAAGGAAAACTGAAGA	
	Sequencing primer		
	SP-CBDA1-2	GAGGAGGCTATGGACC	583, 584, 588, 613, 637
	SP-CBDA2	TGGTTGCTGTCCCAA	758 (3 bp InDel)

Table 3. SNP positions and the percentage of single nucleotide substitution between Type I and III cannabis in *THCAS* gene using pyrosequencing

Name	Chemotype	SNP positions											
		612	**678**	**699**	**744**	**749**	**763**	**862**	**864**	**869**	**881**	**885**	**887**
WT	I	T (86%)	G/A (59:41%)	T/A (57:43%)	G (97%)	C (98%)	T (92%)	G/A (66:34%)	A (86%)	T (82%)	T/G (75:25%)	A/T	A
LOP-1	I	0	G/A (71:29%)	T/A (70:30%)	G (74%)	C (85%)	T (98%)	G/A (78:22%)	A (86%)	T (86%)	T (85%)	A/T	A
LOP-2	I	0	G/A (27:73%)	T/A (28:72%)	0	0	0	G/A (52:48%)	A (84%)	T (86%)	T (76%)	A/T	A
SP	I	T (88%)	G/A (49:51%)	T/A (45:55%)	G (99%)	C (98%)	T (100%)	G/A (72:28%)	A (87%)	T (87%)	T (76%)	A/T	A
PT	I	T (80%)	G/A (45:55%)	T/A (44:56%)	G (91%)	C (91%)	T (92%)	G/A (73:27%)	A (88%)	T (86%)	T (76%)	A/T	A
CHP	I	T (82%)	G (77%)	T (78%)	G (96%)	C (96%)	T (96%)	G (85%)	A (90%)	T (94%)	T (83%)	A/T	A
KRKW	I	T/C (31:69%)	G/A (28:72%)	T/A (27:73%)	G/T (38:62%)	C/T (41:59%)	T/G (38:62%)	G/A (32:68%)	A/G (60:40%)	T/C (51:49%)	T/G (41:59%)	A/T	A/G
KD	I	T (81%)	G/A (64:36%)	T/A (57:43%)	G (96%)	C (100%)	T (86%)	G/A (69:31%)	A (90%)	T (86%)	T (77%)	A/T	A
KK	I	C (82%)	A (77%)	A (78%)	T (81%)	T (84%)	T/G (26:74%)	A (75%)	A/G (70:30%)	T/C (51:49%)	G (80%)	T	A/G
BRR	III	C (77%)	A (76%)	T/A (26:74%)	T (84%)	T (80%)	T/G (27:73%)	A (81%)	A/G (53:47%)	T/C (37:63%)	G (79%)	T	A/G
DNM	III	C (97%)	A (87%)	A (86%)	T (100%)	T (91%)	G (87%)	A (93%)	G (77%)	C (83%)	G (91%)	T	G
CRST	III	C (87%)	A (82%)	A (80%)	T (88%)	T (87%)	G (81%)	A (82%)	A/G (52:48%)	T/C (34:66%)	G (86%)	T	A/G
CHLA	III	C (89%)	A (85%)	A (87%)	T (83%)	T (76%)	T/G (27:73%)	A (78%)	A/G (58:42%)	T/C (38:62%)	T/G (27:73%)	T	A/G
SNP		T→C	G→A	T→A	G→T	C→T	T→G	G→A	A→G	T→C	T→G	A→T	A→G
Amino acid positions		204	226	233	248	250	255	288	288	290	294	295	296
Amino acid change/[1]		NS	NS	NS	NS	A→V	S→A	V→M	V→M	M→T	I→A	NS	K→R
Predicted class/[2]		-	-	-	-	Neutral	Neutral	Neutral	Neutral	Functional	Functional	-	Neutral

[1/] NS stands for nucleotide substitution, but amino acid does not change (synonymous SNP) [2/] dash (-) stands for predicted class does not perform

Table 4. SNP and InDel positions with the percentage of the allele in each locus of *CBDAS* gene using pyrosequencing

Name	Chemotype	SNP positions					
		583	**584**	**588**	**613**	**637**	**758 (3 bp InDel)**
WT	I	A/C (52:48%)	T (89%)	T (88%)	A/G (54:46%)	C/G (40:60%)	C A A (100%)
LOP-1	I	A/C (59:41%)	T (88%)	T (93%)	A/G (56:44%)	C/G (48:52%)	C A A (81%)
LOP-2	I	A/C (57:43%)	T (79%)	T (90%)	A/G (65:35%)	C/G (43:57%)	- - - (67%)
SP	I	A (82%)	T (85%)	C/T (35:65%)	A/G (72:28%)	G (83%)	C A A (89%)
PT	I	A/C (73:27%)	T (82%)	C/T (37:63%)	A/G (65:35%)	G (81%)	C A A (97%)
CHP	I	A/C (35:65%)	T (96%)	T (93%)	A/G (31:69%)	C/G (54:46%)	C A A (92%)
KRKW	I	A/C (58:42%)	T (94%)	T (93%)	A/G (56:44%)	C/G (53:47%)	C A A (92%)
KD	I	A/C (46:54%)	T (87%)	T (89%)	A/G (56:44%)	G (78%)	C A A (100%)
KK	I	A/C (46:54%)	T (83%)	T (93%)	A/G (54:46%)	G (81%)	C A A (100%)
BRR	III	A (85%)	G (91%)	C (89%)	A (82%)	C/G (70:30%)	- - - (95%)
DNM	III	A (97%)	G (91%)	C (99%)	A (96%)	C (86%)	- - - (87%)
CRST	III	A/C (58:42%)	T (86%)	T (91%)	A/G (57:43%)	C/G (34:66%)	C A A (77%)
CHLA	III	A/C (72:28%)	T (93%)	T (82%)	A/G (70:30%)	C/G (63:37%)	C A A (100%)
SNP		A→C	G→T	C→T	A→G	C→G	T, K → -
Amino acid positions		195	195	196	205	213	253, 254
Amino acid change		R→I, L	R→I, L	NS	I→V	H→D	Protein gap
Predicted class/[1,2]		Functional		NP	Neutral	Neutral	NP

[1/] NS stands for nucleotide substitution, but amino acid does not change (synonymous SNP)
[2/] NP stands for predicted class does not perform

Progeny test to verify the highly predictive SNP positions

The SNP percentage of the cross between DNM x CHP and the progeny (F1) at various positions (Table 5.) The results demonstrated that at position 869, DNM (Type III female parent) showed homozygous C alleles at 83%. Whereas CHP (Type I male parent) showed homozygous T alleles at 94%. All ten F1 progeny showed heterozygous T and C alleles (T/C). The percentages of the two alleles are different in each plant, The progeny comprised DCP22-1,

DCP22-2, DCP22-3, DCP22-4, DCP22-5, DCP22-6, DCP22-7, DCP22-8, DCP22-9, and DCP22-10 showed heterozygous alleles as the percentage of T/C allele at 45:55 %, 36:64 %, 71:29%, 41:59%, 55:45%, 47:53%, 55:45%, 31:69%, 32:68%, and 59:41%, respectively. HPLC analysis was performed to confirm the potential of early selection of the molecular markers for the applications of varietal improvement programs in cannabis.

Tetra-primer ARMS-PCR

Tetra-primer ARMS-PCR of *THCAS* gene at position 869 (T/C), the primer was named T869 and designed using the Primer1 program. The sequence of tetra primers and product size of outer and inner primers are shown in Table 6. Eleven cannabis were analyzed by the tetra-primer ARMS-PCR technique, the results of agarose gel electrophoresis are shown in Figure 1. The results illustrated that WT, LOP1, LOP2, SP, PT, and CHP varieties have homozygous T alleles (257 bp) in contrast to the DNM showed C alleles (319 bp). Besides, the band heterozygous alleles (T/C) of F1 progeny (DCP22-1, DCP22-2, DCP22-3, DCP22-4) were detected on agarose gel (257 and 319 bp).

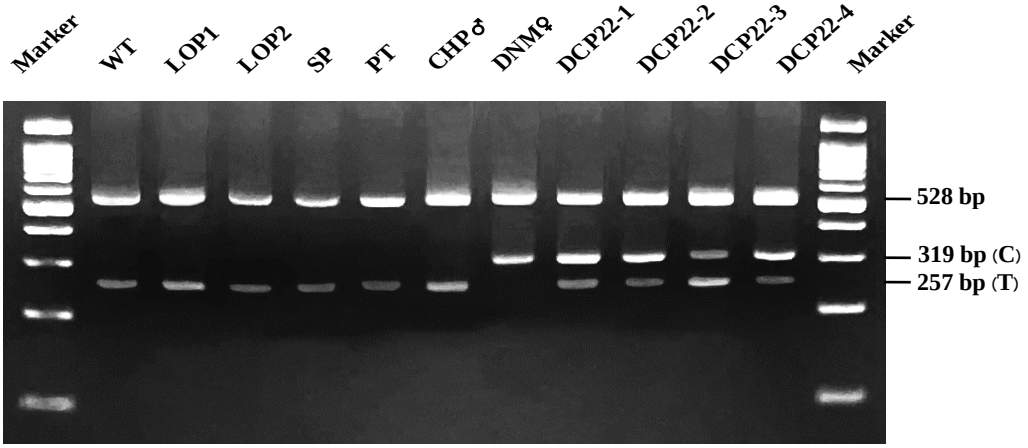

Figure 1. Detection of homozygous and heterozygous alleles at position 869 in *THCAS* gene. 100 bp marker was used in this study. Type I *cannabis* showed T alleles (homozygous). Type III *cannabis* showed C alleles (homozygous). The progeny (DCP22) of DNM x CHP showed heterozygous (T/C)

Table 5. SNP positions and percentage of single nucleotide substitution between DNM x CHP parents and progeny (F1) in *THCAS* gene using pyrosequencing and the cannabinoid content analyzed by HPLC

Name	SNP positions											
	612	**678**	**699**	**744**	**749**	**763**	**862**	**864**	**869**	**881**	**THC %**	**CBD %**
DNM	T (82%)	G (77%)	T (78%)	G (96%)	C (96%)	T (96%)	G (85%)	A (90%)	C (83%)	G (91%)	0.5	9
CHP[1]	C (97%)	A (87%)	A (86%)	T (100%)	T (91%)	G (87%)	A (93%)	G (77%)	T (94%)	T (83%)	ND	ND
DCP22-1	C (86%)	A (75%)	A (82%)	T (76%)	T (76%)	T/G (32:69%)	A (77%)	A/G (74:26%)	T/C (45:55%)	G (79%)	1.85	2.66
DCP22-2	C (79%)	A (83%)	A (86%)	T (92%)	T (79%)	T/G (25:75%)	A (83%)	A/G (61:39%)	T/C (36:64%)	G (78%)	2.51	4.48
DCP22-3	T/C (57:43%)	G/A (40:60%)	T/A (35:65%)	G/T (65:35%)	C/T (72:28%)	T/G (71:29%)	G/A (53:48%)	A (82%)	T/C (71:29%)	T/G (60:40%)	7.43	0.37
DCP22-4	C (86%)	A (84%)	A (82%)	T (92%)	T (82%)	G (79%)	A (78%)	A/G (63:37%)	T/C (41:59%)	T/G (29:71%)	4.30	6.58
DCP22-5	ND	ND	ND	ND	ND	ND	G/A (47:53%)	A/G (52:48%)	T/C (55:45%)	T/G (43:57%)	2.24	5.68
DCP22-6	ND	ND	ND	ND	ND	ND	G/A (32:68%)	A/G (38:62%)	T/C (47:53%)	T/G (42:58%)	1.13	4.86
DCP22-7	ND	ND	ND	ND	ND	ND	G/A (48:52%)	A/G (51:49%)	T/C (55:45%)	T/G (48:52%)	1.70	4.79
DCP22-8	ND	ND	ND	ND	ND	ND	G/A (31:69%)	A/G (38:62%)	T/C (31:69%)	T/G (25:75%)	3.16	6.29
DCP22-9	ND	ND	ND	ND	ND	ND	G/A (25:75%)	A/G (29:71%)	T/C (32:68%)	T/G (16:84%)	2.22	4.91
DCP22-10	ND	ND	ND	ND	ND	ND	G/A (45:55%)	A/G (55:45%)	T/C (59:41%)	T/G (32:68%)	2.06	4.75

[1]/ CHP is a male plant of chemotype I (high THC) cannabis. ND stands for does not determine

Table 6. Oligonucleotide sequence of tetra-primer ARMS-PCR for detection of SNP position 869 in *THCAS* gene

Name	Primer Sequence (5'-3')	Product size (bp)	Target allele[1]
Forward outer	ATGTTGATGGAAAAGTTCTAGAT	528	
Reverse outer	AGTCTAACTTAATTGAGAAAGCC		
Forward inner	ACAAGTATGACAAAGATTTAGTACTAAC	319	C
Reverse inner	CTTTGTTATGAAGTGAGGCA	257	T

[1/]Specific allele at position 869: C allele represents Type III cannabis. T allele represents Type I cannabis.

Discussion

The previous study reported that the phytocannabinoids of cannabis were synthesized from precursor cannabigerolic acid (CBGA) by three different enzymes. CBGA is metabolized into tetrahydrocanabinolic acid (THCA) via THCA synthase, into cannabidiolic acid (CBDA) via CBDA synthase, and this research concentrated on *THCA* and *CBDA* synthase (*THCAS* and *CBDAS*) genes. The *THCAS* and *CBDAS* nucleotide sequences discussed here are full-length of 1635 bp and encode for 545 and 544 amino acids, respectively (Onofri *et al.,* 2015). The expression of these genes appears to be the major factor effect to accumulation of cannabinoid content. Nevertheless, the mechanisms regulating the functions of these genes need to be researched more. In this study, the nucleotide sequences of *THCAS* and *CBDAS* genes were derived from high-THC cannabis (Type I) and high-CBD cannabis (Type III). The sequences were aligned and compared separately to find the SNPs associated with the cannabinoid type and composition of cannabinoids from the maturity inflorescence of the cannabis plant. Of the 12 SNP loci detected by pyrosequencing, two loci at positions 869 (T/C) and 881 (T/G) were found to have the single nucleotide mutation effect of triplet code translate to the functional amino acid positions 290 (M→T) and 294 (I→A), respectively.

To verify the highly predictive SNP position 869 by progeny analysis. The cross was made between a female DMN plant (a homozygous CBD dominant plant) and a male CHP plant (a homozygous THC dominant plant). The resulting F1 genotypes at SNP position 869 were heterozygous and contained two different alleles (T/C). However, there are six F1 plants that have genotypes containing an almost equal ratio of T and C alleles. In addition, the T allele at position 869 on *THCAS* gene was the expected allele related to the *THCAS* gene. Furthermore, the C allele was the expected allele associated with the *CBDAS* gene. These alleles correspond with the

percentage of Type III plant alleles containing a higher percentage of C allele than T allele.

T869 marker could detect homozygous and heterozygous genotypes in cannabis, which could be used for the early selection in the varietal improvement of cannabis. In addition, the homozygous allele could be used to predict the cannabinoid type correctly but not the quantity. At the same time, we could use the percentage of heterozygous alleles to predict the trend of cannabinoid type in the progeny.

The phenotype of quantitative cannabinoid content showed that all F1 plants produce both THC and CBD, but CBD content was higher than THC content except for DCP22-3. At this point, we can conclude that the inheritance of the *CBDAS* gene was dominant over the *THCAS* gene, and the genetic inheritance involved in regulating cannabinoids production are quantitative trait loci (QTL) which mean it was regulated by many minor genes or gene family and its expression can also be affected by an environment. However, recent research has found genetic support that there are multiple genes in close proximity responsible for the production of cannabinoids (Vergara *et al*., 2019).

Acknowledgements

The author would like to express our appreciation to the Cannabis Foundation of Rangsit University for the partial financial support and the College of Agricultural Innovation and Food Technology, Rangsit University, Thailand. We also thank Medical Cannabis Research Institute College of Pharmacy and High Generation Technology CO., LTD. For quantification of cannabinoid content by HPLC.

References

Brooks, L. D. (2003). Single Nucleotide Polymorphisms. In P.-Y. Kwok (Ed.), Single Nucleotide Polymorphisms Methods and Protocols, Totowa, New Jersey: Humana Press Inc. 212:1-4.

Cascini, F., Farcomeni, A., Migliorini, D., Baldassarri, L., Boschi, I., Martello, S. and Bernardi, J. (2019). Highly Predictive Genetic Markers Distinguish Drug-Type from Fiber-Type Cannabis sativa L. Plants (Basel, Switzerland*)*, 8*:*496.

Doyle, J. J. and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. PHYTOCHEMICAL BULLETIN, 19*:*11-15.

Fakruddin, M. and Chowdhury, A. (2012). Pyrosequencing-An Alternative to Traditional Sanger Sequencing. American Journal of Biochemistry and Biotechnology, 8*(1)*. doi:10.3844/ajbbsp.2012.14.20

Garfinkel, A. R., Otten, M. and Crawford, S. (2021). SNP in Potentially Defunct Tetrahydrocannabinolic Acid Synthase Is a Marker for Cannabigerolic Acid Dominance in Cannabis sativa L. Genes, 12:228.

Kojoma, M., Seki, H., Yoshida, S. and Muranaka, T. (2006). DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in “drug-type” and “fiber-type” Cannabis sativa L. Forensic Science International, 159*:*132-140.

Matsuda, K. (2017). Chapter Two - PCR-Based Detection Methods for Single-Nucleotide Polymorphism or Mutation: Real-Time PCR and Its Substantial Contribution

Toward Technological Refinement. In G. S. Makowski (Ed.), Advances in Clinical Chemistry, 80:45-72.
Onofri, C., de Meijer, E. P. M. and Mandolino, G. (2015). Sequence heterogeneity of cannabidiolic- and tetrahydrocannabinolic acid-synthase in Cannabis sativa L. and its relationship with chemical phenotype. Phytochemistry, 116:57-68.
Posselt, U. (2010). Breeding Methods in Cross-Pollinated Species, pp.39-87.
Ramirez, C. L., Fanovich, M. A. and Churio, M. S. (2019). Chapter 4 - Cannabinoids: Extraction Methods, Analysis, and Physicochemical Characterization. In R. Atta ur (Ed.), Studies in Natural Products Chemistry, 61:143-173.
Ranalli, P. (2004). Current status and future scenarios of hemp breeding. Euphytica, 140:121-131.
Salentijn, E. M. J., Zhang, Q., Amaducci, S., Yang, M. and Trindade, L. M. (2015). New developments in fiber hemp (*Cannabis sativa* L.) breeding. Industrial Crops and Products, 68:32-41.
Schilling, S., Dowling, C., Shi, J., Ryan, L., Hunt, D., Oreilly, E. and Melzer, R. (2020). The Cream of the Crop: Biology, Breeding and Applications of Cannabis sativa.
Singh, D. P., Singh, A. K., and Singh, A. (2021). Chapter 1 - Plant breeding: past, present, and future perspectives. In D. P. Singh, A. K. Singh, & A. Singh (Eds.), Plant Breeding and Cultivar Development, Academic Press, pp.1-24.
Sirikantaramas, S. and Taura, F. (2017). Cannabinoids: Biosynthesis and Biotechnological Applications.
Sundru, M., Archer, A. and Halami, P. M. (2015). Screening, Characterization and In Vitro Evaluation of Probiotic Properties Among Lactic Acid Bacteria Through Comparative Analysis. Probiotics and antimicrobial proteins, 7. doi:10.1007/s12602-015-9195-5.
Vergara, D., Huscher, E., Keepers, K., Givens, R., Cizek, C., Torres, A. and Kane, N. (2019). Gene copy number is associated with phytochemistry in Cannabis sativa. AoB PLANTS, 11, plz074. doi:10.1093/aobpla/plz

(Received: 8 October 2022, accepted: 30 April 2023)