# Control genético del quimiotipo en *Cannabis sativa*

# Genetic control of the chemotype in *Cannabis sativa*

A. Gila-Navarro*, J. Weiss, M. Egea-Cortines

Instituto de Biotecnología Vegetal, Departamento de Ingeniería Agronómica, Universidad Politécnica de Cartagena. Campus Muralla de Mar, s/n, edificio I+D+i, Cartagena 30202, Murcia, España.

*alberto.gila@upct.es

***Resumen***

***El modelo actual de herencia del quimiotipo en Cannabis sativa se centra en las dos sintasas terminales de la ruta biosintética de los cannabinoides, THCAS y CBDAS. Sin embargo, el modelo ignora la posible contribución de la regulación de la expresión génica de estas enzimas a la determinación del quimiotipo. En el presente trabajo en curso se muestra una red de regulación génica con los posibles factores de transcripción que regulan a THCAS y CBDAS y se debaten posibles implicaciones.***

**Palabras clave:** Cannabinoides; expresión génica; factores de transcripción; red de regulación génica.

***Abstract***

***The current model of chemotype inheritance in Cannabis sativa revolves around the two terminal synthases of the cannabinoid biosynthesis pathway, the THCAS and CBDAS. However, the model ignores the possible contribution of the regulation of these enzymes' gene expression to the determination of the chemotype. In the present on-going work, we present a gene regulatory network inferred in silico with the candidate transcription factors that regulate the expression of the THCAS and CBDAS and discuss possible implications.***

**Keywords**: Cannabinoids; gene expression; transcription factors; gene regulatory networks.

## 1. INTRODUCTION

Cannabis is today the source of a multibillion-dollar industry pharmaceutical industry thanks to the compounds it produces, the cannabinoids. Specifically, THC (Δ-9 tetrahydrocannabinol) and CBD (cannabidiol) have therapeutic value in treating insomnia, chronic pain, and cancer among other applications. (1), (2), (3).

Therefore, the single most relevant trait of the plant is the synthesis of cannabinoids. Different varieties have a different quantity and relative abundance of these compounds. The final step of the pathway is where the genetic model focuses on. Here, several terminal synthases, like THCAS and CBDAS compete for a common precursor called CBGA to form different cannabinoids.(4)

The model is a biallelic codominant one with one locus for the THCAS and one for the CBDAS. These loci are comprised of several paralogs tightly linked that get inherited together, but only one of them is actively expressed. The model divides the alleles of THCAS and CBDAS into functional and non-functional. Depending on the combination of alleles, the plant will produce a different ratio of THC to CBD. A plant with two, one of none functional allele for a particular synthase will accumulate high, low or null quantities of that cannabinoid .(5)

The model however does not consider the likely effect of gene regulation of the synthases on the chemotype determination. This would be an independent extra layer of control on top of the alleles. Mutations in either CIS and/or TRANS regulatory elements like promoters and

transcription factors may also contribute to the chemotype. In this work we present a gene regulatory network inferred *in silico* with the putative candidates that regulate the expression of the CBDAS and THCAS. Understanding the control of the expression of the cannabinoid synthases will provide insight into other elements that contribute to the chemotype and may be targets for breeders.

## 2. MATERIALS AND METHODS

### 2.1 Data used

We downloaded from the NCBI database a total of 123 RNAseq datasets from *Cannabis sativa* trichomes. The data originates from 4 different studies (NCBI BioProject accession numbers: PRJNA560453, PRJNA599437, PRJNA498707, PRJNA483805) and contains a total of 16 varieties. This data was combined with 15 additional RNAseq dataset generated by ourselves originating from 3 varieties of the company LinneoHealth S.L.

### 2.2 Differential gene expression

The data in FASTQ format was first analyzed with FastQC and the reads were edited when necessary with Trimmomatic. To quantify the gene expression, we used Salmon and the R package Tximport from Bioconductor. Lowly expressed genes were discarded. Two negative-binomial GLMs (generalized linear model) were fit with the edgeR package. The independent variables are the *variety* and the *experiment* from which it comes. The model using the *variety* was used to perform all 1-vs-1 and 1-vs-all comparisons between plants. The model using the *experiment* was used to mitigate the batch effect.

### 2.3 Dataset visualization

The count matrix was transformed to log2CPM using the normalization factors calculated with the TMM method of edgeR. The PCA method was used as dimensionality reduction. The expression level of the cannabinoid synthases was then visualized as a barplot.

### 2.4 Gene regulatory networks

A list of TFs (Transcription Factors) from *Cannabis sativa* was retrieved after annotating with GO terms the reference transcriptome using Sma3. The keywords “transcription factor” were used to obtain a list of possible TFs. This was then manually curated and filtered to include only differentially expressed TFs in any of the contrasts. The end list contains 870 genes.

These TFs together with the log2CPM transformed count matrix were inputed to ARACNe to infer the gene regulatory network.

## 3. RESULTS AND DISCUSSION

The expression pattern of the THCAS and CBDAS enzymes differs. As can be seen in Fig.1 , the THCAS is expressed moderately in many varieties. However, CBDAS is highly expressed only in some varieties. The varieties with an above average level of expression are coincidentally the ones that accumulate a high quantity of CBD. This indicates that the CBDAS expression could be used as a marker to discriminate high and low CBD producing chemotypes. The production of CBD is thus determined not only by the allele a plant has, like it has been proven in (5) and (6) but also by its expression level. We do not know yet if this is due to CIS or TRANS regulatory elements polymorphisms.

For the THCAS, gene expression alone cannot explain the differences in chemotype. Other factors such as sequence polymorphisms that affect the catalytic activity or the enzyme localization within the cell may play a bigger role in this case. Other studies have identified QTLs related to CBD/THC balance which are located in different chromosomes and may be related to this.(7)

The gene regulatory network (Fig.2) shows that CBDAS and THCAS are mostly controlled by a different set of TFs. CBDAS was also linked to fewer TFs than THCAS. Interestingly, the TFs that regulate CBDAS according to the network include two transposases (RICESLEEPER and DAYSLEEPER)(8). The implications of this links are unknown. Regarding THCAS, it is controlled by TFs belonging to families that in *Arabidopsis thaliana* are known to be related to abiotic stress response, like SRM, GT, RAP or NLP

There are two studies that identified TFs from *Cannabis sativa* involved in cannabinoid (9) (10) and anthocyanin biosynthesis belonging to the AP2, MYB and WRKY families. The MYB and AP2 family are represented in our network but the TFs from these works do not appear. An experimental validation of our network through yeast-one-hybrid assay is currently undergoing

## 4. CONCLUSSIONS

THCAS expression is not directly correlated to the synthesis of THC. CBDAS expression on the contrary can discriminate between high and low CBD-producing plants. The expression of the synthases is controlled by a different set of TFs which are related to different processes.

## 5. AKNWOLEDGEMENTS

This work was funded by Project PID2021-127933OB-C21 financed by MCIN/AEI/10.13039/501100011033 and FEDER to M. Egea-Cortines and J. Weiss. A. Gila-Navarro received an FPU fellowship FPU21/02130 from the Spanish Ministry of Universities

## 6. REFERENCES

1. Kuhathasan N, Dufort A, MacKillop J, Gottschalk R, Minuzzi L, Frey BN. The use of cannabinoids for sleep: A critical review on clinical trials. Exp Clin Psychopharmacol. 2019;27:383–401.

2. Johal H, Devji T, Chang Y, Simone J, Vannabouathong C, Bhandari M. Cannabinoids in Chronic Non-Cancer Pain: A Systematic Review and Meta-Analysis. Clin Med Insights Arthritis Musculoskelet Disord. 2020 Jan;13:117954412090646.

3. Mangal N, Erridge S, Habib N, Sadanandam A, Reebye V, Sodergren MH. Cannabinoids in the landscape of cancer. J Cancer Res Clin Oncol. 2021 Sep 1;147(9):2507–34.

4. 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;230(1):73–89.

5. Wenger JP, Dabney III CJ, ElSohly MA, Chandra S, Radwan MM, Majumdar CG, et al. Validating a predictive model of cannabinoid inheritance with feral, clinical, and industrial Cannabis sativa. Am J Bot. 2020;107(10):1423–32.

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

7. Campbell LG, Dufresne J, Sabatinos SA. Cannabinoid Inheritance Relies on Complex Genetic Architecture. Cannabis Cannabinoid Res. 2020 Mar;5(1):105–16.

8. Knip M, de Pater S, Hooykaas PJ. The SLEEPERgenes: a transposase-derived angiosperm-specific gene family. BMC Plant Biol. 2012 Oct 16;12(1):192.

9. 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 1;106(1):49–65.

10. Kundan M, Gani U, Fayaz Mohd, Angmo T, Kesari R, Rahul VP, et al. Two R2R3-MYB transcription factors, CsMYB33 and CsMYB78 are involved in the regulation of anthocyanin biosynthesis in Cannabis sativa L. Ind Crops Prod. 2022 Nov 15;188:115546.

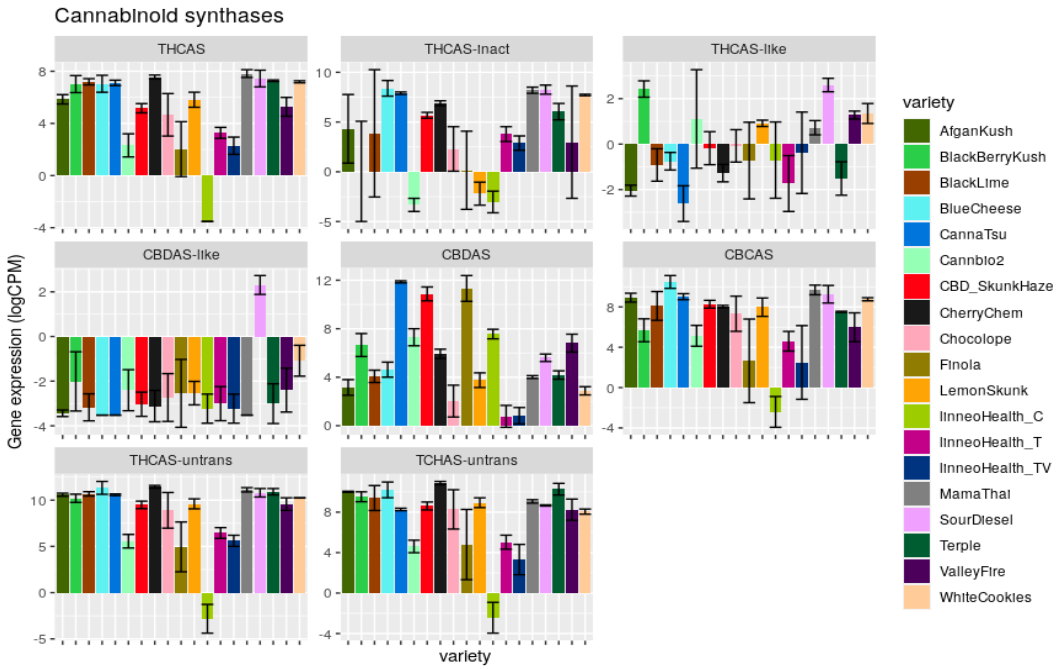


**Figure 1**. Cannabinoid synthases gene expression in logCPM values

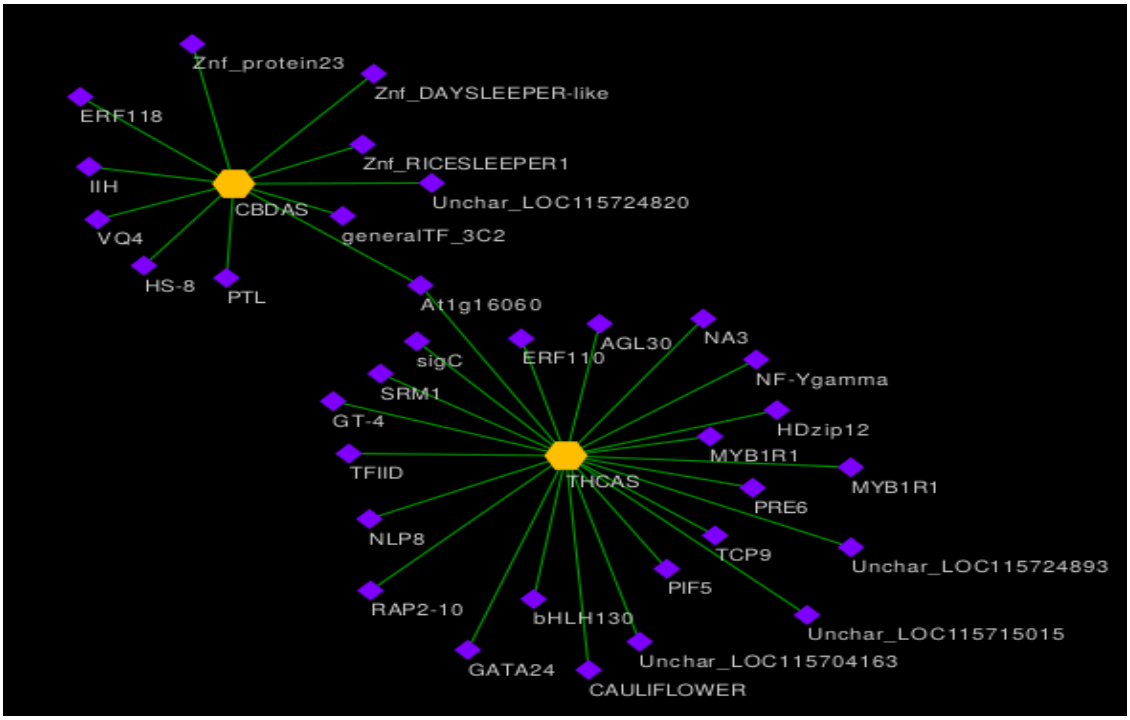


**Figure 2**: ARACNe gene regulatory network of CBDAS and THCAS