



Transcriptomic dissection of floral identity and trichome-derived metabolite pathways in *Cannabis sativa* L.

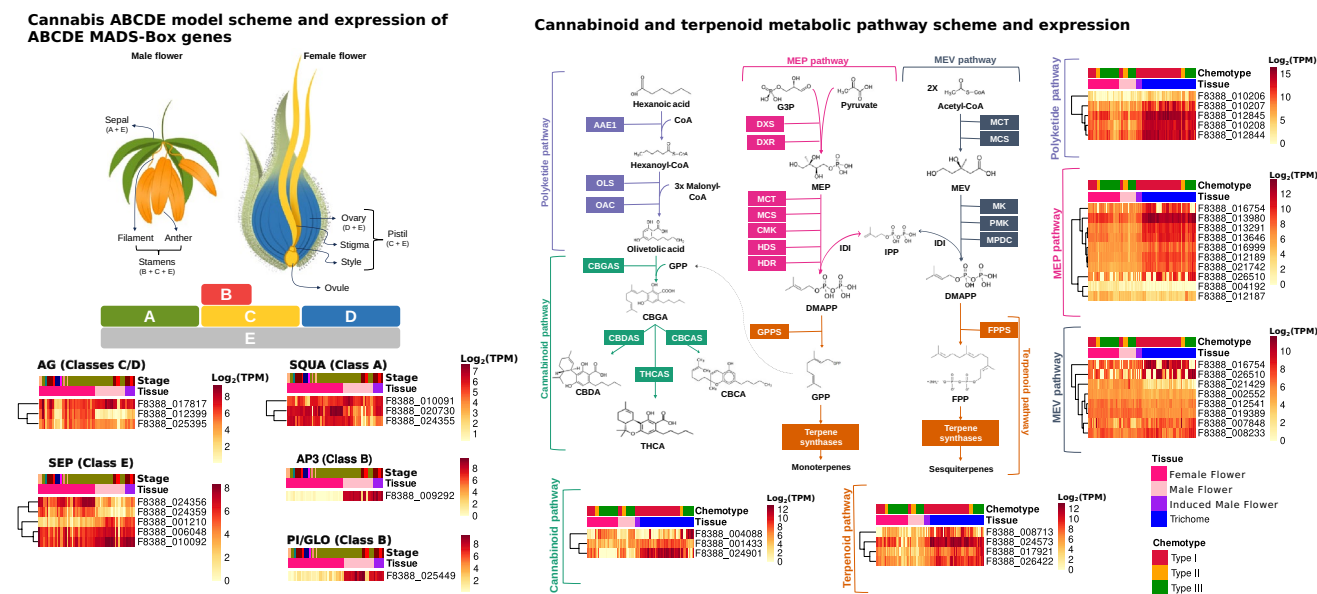
Kevelin Barbosa-Xavier¹ · Thiago M. Venancio¹

Received: 9 March 2026 / Accepted: 30 May 2026
 © The Author(s) 2026

Abstract

Cannabis sativa L. has a long history of medicinal and industrial use, with female flowers as the primary source of bioactive cannabinoids and terpenoids synthesized in glandular trichomes. Despite its importance, the genetic mechanisms governing flower development, sex determination, and secondary metabolite biosynthesis remain incompletely understood. Here, we combined transcriptomic and comparative genomic analysis to elucidate the molecular networks underlying these traits. Integrating 117 RNA-Seq datasets and phylogenetic analyses across nine *C. sativa* genomes, we identified 31 orthogroups of MADS-box transcription factors. Expression profiling highlighted distinct candidates for male and female flower identity, consistent with the ABCDE model. AP3, PI/GLO, and MIKCS clades were preferentially expressed in male flowers, whereas AGL6, FLC-like, and Bsister genes predominated in female flowers. Analyses of pollen-associated genes underscored roles for sugar metabolism and transport in male fertility. Profiling cannabinoid and terpenoid biosynthetic genes confirmed strong trichome expression and revealed a key distinction: upstream polyketide, MEV, MEP, and the terpenoid pathways showed conserved expression across chemotypes, whereas the cannabinoid pathway displayed chemotype-specific profiles. Collectively, these findings provide an integrative framework for floral development and secondary metabolism in *C. sativa*, offering targets for functional validation and breeding to optimize agronomic and medicinal traits in modern cultivars.

Graphic abstract



Keywords ABCDE model · Flower development · Cannabinoids · Photoperiod response · Terpenoids · MADS-box genes

Communicated by Stefan de Folter.

Extended author information available on the last page of the article

Introduction

The use of *Cannabis sativa* L. female flowers (FF) dates back over 2,700 years, serving as medicinal, ritualistic, and recreational purposes across diverse cultures (Ren et al. 2019). In modern contexts, cannabis is primarily consumed via two methods: combustion (smoking) and essential oil extraction. FF are the most valuable part of the plant due to their rich biochemical profile (Pieracci et al. 2021; Ren et al. 2019). These flowers are densely covered in glandular trichomes—microscopic resin-producing structures that synthesize a wide array of secondary metabolites, including cannabinoids, terpenoids, and flavonoids. These compounds are responsible for the plant therapeutic, psychoactive, and sensory properties (ElSohly & Slade 2005; Ren et al. 2019).

The primary metabolites produced in cannabis FF are cannabinoids, hydrocarbons, sugars, steroids, mono- and sesquiterpenes, flavonoids, amino acids, and nitrogenous compounds (ElSohly & Slade 2005). Among these, cannabinoids are the most pharmacologically and commercially valuable, particularly delta-9-tetrahydrocannabinol (Δ^9 -THC) and cannabidiol (CBD), which are associated with both medicinal and adult-use applications (Gülck & Møller 2020; Leinen et al. 2023). Terpenoids also contribute significantly, especially in the context of adult use, as they shape the aroma and flavor profile of cannabis products (Hanuš & Hod 2020; Rice & Koziel 2015). Notably, the effects observed from smoking cannabis or using full-spectrum oils result from the combined action of multiple metabolites produced in trichomes, a phenomenon known as “entourage effect” (Ferber et al. 2020).

C. sativa is a dioecious species, producing separate male and female plants. Although studies have shown that flower sex can be altered by chemical or environmental cues (Adal et al. 2021), the genetic mechanisms underlying flower development and sex differentiation remain largely

unresolved. In this study, we leveraged a comprehensive collection of RNA-Seq datasets from the Cannabis Expression Atlas (CannAtlas) (Barbosa-Xavier et al. 2024) to investigate three key gene groups involved in cannabis flower biology: (i) MADS-box transcription factors (TFs); (ii) genes potentially associated with pollen development; and (iii) genes involved in cannabinoid and terpenoid biosynthetic pathways. These transcriptomic analyses were complemented by comparative genomics across multiple *Cannabis* genomes. Our findings reveal candidate genes associated with flower morphology and sex determination, as well as key enzymes driving cannabinoid and terpenoid production in trichomes.

Materials and methods

Expression data

Expression data (in transcripts per million, TPM) from 117 RNA-Seq samples—comprising trichomes ($n=59$), FF ($n=34$), male flowers (MF, $n=18$), and induced male flowers (IMF, $n=6$)—along with their metadata, were obtained from the CannAtlas (Barbosa-Xavier et al. 2024).

MADS-box phylogenetic and orthology analysis

We used the PlantTFDB v5.0 prediction tool (Tian et al. 2020) to identify MADS-box TFs across nine *C. sativa* varieties with publicly available genome and protein sequence data (Table 1). To determine orthologous relationships across these nine *C. sativa* genomes, we employed OrthoFinder v3.0.1b1 (Emms & Kelly 2019). The MADS-box protein sequences reported by Ristevski (2023) were used as references to identify orthogroups (OGs) containing MADS-box genes. Only OGs with at least one reference MADS-box protein were retained for downstream analysis. Protein sequences within each OG were aligned using

Table 1 Cannabis varieties used in MADS-box analysis

Genome	Reference	Source database Protein sequence source
Jamaican Lion mother + Y	(K. McKernan et al. 2018, 2020)	RefSeq*
Pink Pepper	(Ryu et al. 2024)	RefSeq
Purple Kush	(Bakel et al. 2011; Lavery et al. 2019)	CannabisGDB
Finola	(Bakel et al. 2011; Lavery et al. 2019)	CannabisGDB
CBDx18 (Cs10)	(Grassa et al. 2018, 2021)	CannabisGDB
LA Confidential	Steep Hill Genetics	CannabisGDB
Chemdog91	Courtage Life Sciences	CannabisGDB
Pineapple Banana Bubba Kush	Steep Hill Genetics	CannabisGDB
Cannatonic	Phylos Bioscience	CannabisGDB

*The Y region was manually collected from Jamaican Lion father files and appended to the Jamaican Lion mother genome (as used in CannAtlas).

MAFFT v7.525 (Katoh & Standley 2013), and maximum-likelihood phylogenies were reconstructed using IQ-TREE v2.3.6 (Minh et al. 2020) under default parameters. Phylogenetic trees were visualized with FigTree v1.4.4 (Rambaut 2012), and OG alignments were further inspected using the NCBI Multiple Sequence Alignment Viewer v1.26.0.

Male flower development genes

We conducted BLAST (Altschul et al. 1997) searches against CannAtlas using three reference gene sets to identify genes potentially related to pollen development and male function in *C. sativa*: (i) 62 sugar metabolism-related reference genes from Liu et al. (2021a, b), (ii) 61 candidate genes associated with masculinization in cannabis proposed by Adal et al. (2021), and (iii) 185 reference genes implicated in male sterility in Chinese cabbage (*Brassica rapa*) (Huang et al. 2020). BLAST searches were performed with the following parameters: *similarity_matrix* = BLOSUM62, *e-value* ≤ 0.001 , *qcov_hsp_perc* ≥ 80 and, *max_target_seqs* = 1. The assignment of the functional category was based on the classification established in each reference study. We used the expression classification from CannAtlas to identify the homologous genes Tissue-Specific to MF or IMF and also the Group-Enriched ones for MF + IMF.

Cannabinoid and terpenoid pathway genes

Genes involved in cannabinoid and terpenoid biosynthetic pathways, as reported by Kovalchuk et al. (2020), were retrieved from Genbank and used to extract expression data from CannAtlas. All genes were successfully identified in CannAtlas. BLAST searches were conducted with the following parameters: *per_identity* = 80, *e-value* $\leq 1e-10$, and *qcov_hsp_perc* ≥ 90 .

Enrichment analysis

Enrichment analyses of Gene Ontology (OG), KEGG pathways, and TF family were performed using a Fisher's exact test in R, with significance defined by a Benjamini–Hochberg adjusted p-value < 0.05 .

Results and discussion

Identification of MADS-box orthogroups across cannabis varieties and analysis of their expression in male and female flowers

The MADS-box TF family plays a central role in the ABCDE genetic model of flower identity, which governs

Table 2 MADS-box TF distribution across nine cannabis varieties

Variety	# Total TFs	# MADS TFs	% MADS
Jamaican Lion mother + Y	1532	70	4.57
Pink Pepper	2315	123	5.31
Purple Kush	1816	77	4.24
Finola	1694	77	4.55
CBDrx18 (Cs10)	1399	76	5.43
LA Confidential	1055	43	4.08
Chemdog91	789	35	4.44
Pineapple Banana Bubba Kush	920	36	3.91
Cannatonic	949	44	4.64
Mean	1385.44	64.56	4.57

the specification and differentiation of sepals, petals, stamens, carpels, and ovules (F. Chen et al. 2017). This model describes how the combinatorial expression of specific MADS-box TFs defines floral organ identity and coordinates flower morphogenesis, ensuring proper reproductive development (Coen & Meyerowitz 1991). Using *Arabidopsis thaliana* MADS-box TFs as reference, Ristevski (2023) identified 68 cannabis MADS-box TFs in the *C. sativa* ‘Cs10’ genome (CBDrx-18), which were distributed across 18 phylogenetic clades. These include members of both type I (M α , M β and M γ) and type II (MIKCS, Bsister, PI/GLO, AP3, AGL12, SVP, AGL15, AGL17, AG, TM3, FLC-like, FLC, AGL6, SQUA, and SEP) MADS-box genes.

To expand this analysis, we used the PlantTFDB prediction tool to identify MADS-box TFs across nine *C. sativa* varieties with publicly available genomes (Table 2). Based on the total number of predicted TFs per genome, we found that MADS-box genes account for an average of 4.57% of all cannabis TFs (Table 2). In the CBDrx-18 (Cs10) genome, our approach identified 76 MADS-box genes, including all the 68 genes reported by Ristevski (2023). This result supports the accuracy and sensitivity of the strategy used in both studies.

OrthoFinder grouped 95.5% of the input genes (280,293 out of 293,409 predicted proteins) into 17,972 OGs. Among these, 0.5% (1,467 genes) were assigned to 596 variety-specific OGs (Table S1), with the majority concentrated in Purple Kush, Jamaican Lion, and Finola, which together accounted for 420 of the variety-specific OGs (Table S2). To explore MADS-box orthology across *C. sativa* varieties, we used OrthoFinder with the 68 MADS-box TFs identified by Ristevski (2023) in the CBDrx-18 (Cs10) genome as reference. We identified 31 OGs containing at least one MADS-box gene, of which 23 included at least one of the reference MADS-box TFs (Table S3). Among these, only two OGs grouped genes from more than one phylogenetic clade: OG0000096 included representatives of AG, AGL6,



Table 3 Number of MADS-box* and non-MADS** genes in each reference clade of each OG

OG	MADS ref. clade	Ref. MADS	CBDRx-18		Cannatonic		Chemdog91		Finola		JL_Mother_Y		LA Confidential		Pineapple BBK		Purple Kush		Pink Pepper		Total
			MADS	Non-MADS	MADS	Non-MADS	MADS	Non-MADS	MADS	Non-MADS	MADS	Non-MADS	MADS	Non-MADS	MADS	Non-MADS	MADS	Non-MADS	MADS	Non-MADS	
OG0000033	AP3	1	1				1				1		1		1		1		1		8
OG0000096	AG	2	3		2		1	1	4	2	2	1	1	1	3		2	1	7		33
OG0000096	AGL6	2	1			1		1	2		1		1	1	1		2		1		14
OG0000096	FLC	1	1	1		1				1	1		1				1		2		10
OG0000096	SEP	9	5	2	5	1		2	7	2	6		1	3	3	3	6	4	19	1	79
OG0000207	AGL17	9	1	4	1	2		2	6	1	5	4	1	2	3	3	1	7	13	4	69
OG0000366	Mg	4	5		5		3		9		3		5		1	2	6		6		49
OG0000433	TM3	3	3	1	1	1	1	1	2		2	1		3	1	1	3		10		34
OG0000433	FLC-like	1	2				1		2		2			1					1		10
OG0000580	Ma	2	2		2	1	2	1	4	2	3	2	2	1	2	2	3	3	5		39
OG0000736	Mb	1	1		1	1	1				1		1			1		1			8
OG0000955	SVP	4	2		1	1	1	1	3		3		2		3		3		6		30
OG0001105	MIKCS	3	1	1		1				2	1	4		3	1	1	1		3	1	24
OG0001107	Mb	2																	3		5
OG0001251	SQUA	3	2		1	1		1	4		3			2	3		3		4		27
OG0001495	MIKCS	2	1		1			1	2		2		1		1				8		19
OG0001697	Ma	3	2				2		3		1		1				4		3		19
OG0001926	Mb	3	3		3				1		1		3		2		2		3		21
OG0001966	Mg	1	2		1		1		5		2		1		1		1		2		16
OG0002119	Ma	2	1		2				2		3		1		1		1		5		18
OG0002194	Bsister	3	1		1		1		2		1	1	1				2		3		16
OG0002853	Ma	2	1		2				1		1		1				1		4		13
OG0003904	Ma	1	1														1				3
OG0005875	PI/GLO	1	1				1			1	1						2				7
OG0006023	AGL12	1	1		1			1		2	1		1		1		1		1		11
OG0007459	AGL15	1					1						1				2				5
OG0008462	Mg	1																	1		2
Total		68	44	9	29	11	17	12	59	13	46	14	24	20	27	12	49	17	112	6	589

*MADS-box genes predicted by PlantTFDB

** Genes belonging to MADS-box orthogroups that have lost the MADS-box domains

a) OG0001926 - Clade Mb

NCBI Multiple Sequence Alignment Viewer, Version 1.26.0

Sequence ID	Start	Alignment	End	Identity	Coverage	Mismatches
Risteveki_c10_MADS_REF XP_030508932.2	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	254	100.00	100.00	0
CBDx18_ref_p_CaCDB_070010520	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	254	100.00	100.00	0
PurpleKush_ref_p_CaPK_000005590	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	254	100.00	100.00	0
PBBKush_ref_p_CaPBB_000147010	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	254	100.00	100.00	0
LAConfidential_ref_p_CaLAC_000013240	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	254	100.00	100.00	0
Cannatonic_ref_p_CaCAN_000013670	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	254	100.00	100.00	0
Finola_ref_p_CaFN_000003390	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	254	99.61	100.00	1
pink_pepper_ref_p_CaPP_000008977.2	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	254	99.61	100.00	1
Risteveki_c10_MADS_REF XP_030508972.2	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	254	99.61	100.00	1
Cannatonic_ref_p_CaCAN_000002530	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	262	91.22	100.00	15
CBDx18_ref_p_CaCDB_060002720	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	262	91.22	100.00	15
PBBKush_ref_p_CaPBB_000007780	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	262	91.22	100.00	15
pink_pepper_ref_p_CaPP_03047446.1	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	262	91.22	100.00	15
PurpleKush_ref_p_CaPK_000001640	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	262	91.22	100.00	15
Risteveki_c10_MADS_REF XP_03047446.1	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	262	91.22	100.00	15
Finola_ref_p_CaFN_000007126.1	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	263	90.87	100.00	15
LAConfidential_ref_p_CaLAC_000019640	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	263	90.87	100.00	15
CBDx18_ref_p_CaCDB_070010900	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	266	90.79	88.98	9
LAConfidential_ref_p_CaLAC_000007010	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	263	80.15	88.58	15
CBDx18_ref_p_CaCDB_060002780	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	233	80.15	88.58	15
PurpleKush_ref_p_CaPK_000001630	1	MLPLKLI INEAKRRTFSKRSLSL KAVELSTLDCVAVLI I KSTNPSNNP AVTPHPRVSLI NDH AKTTPSSKSLNTD LTKERN NVNMEI EKLCSQESKGV VVNGSKERL NQGLCFASELDMKQLNRHRAI SSMHKLQALLOCHOM SSSSSAP	233	79.77	88.58	16

b) OG0000096 - Clade AGL6

NCBI Multiple Sequence Alignment Viewer, Version 1.26.0

Sequence ID	Start	Alignment	End	Identity	Coverage	Mismatches
Risteveki_c10_MADS_REF XP_03050965.2	1	MGKGVLEIR ENI INQVFSKRRNLKAVELHMLDCAEVL I FSSRQLYFEGSSNTI VERRYSTALONNVEGEPNLI LEI PLKAKYVES OTTERPHENIEELGKELQNEKLDRTLSOAHOMVADLLEI RONDHRETNELQKSLKEENKGRISA	244	100.00	100.00	0
CBDx18_ref_p_CaCDB_050001490	1	MGKGVLEIR ENI INQVFSKRRNLKAVELHMLDCAEVL I FSSRQLYFEGSSNTI VERRYSTALONNVEGEPNLI LEI PLKAKYVES OTTERPHENIEELGKELQNEKLDRTLSOAHOMVADLLEI RONDHRETNELQKSLKEENKGRISA	225	93.78	90.98	11
Cannatonic_ref_p_CaCAN_000014960	1	MGKGVLEIR ENI INQVFSKRRNLKAVELHMLDCAEVL I FSSRQLYFEGSSNTI VERRYSTALONNVEGEPNLI LEI PLKAKYVES OTTERPHENIEELGKELQNEKLDRTLSOAHOMVADLLEI RONDHRETNELQKSLKEENKGRISA	82	62.93	32.38	11
Chemdog91_ref_p_CaCD91_000002060	1	MGKGVLEIR ENI INQVFSKRRNLKAVELHMLDCAEVL I FSSRQLYFEGSSNTI VERRYSTALONNVEGEPNLI LEI PLKAKYVES OTTERPHENIEELGKELQNEKLDRTLSOAHOMVADLLEI RONDHRETNELQKSLKEENKGRISA	74	89.19	29.10	5
Finola_ref_p_CaFN_010001010	1	MGKGVLEIR ENI INQVFSKRRNLKAVELHMLDCAEVL I FSSRQLYFEGSSNTI VERRYSTALONNVEGEPNLI LEI PLKAKYVES OTTERPHENIEELGKELQNEKLDRTLSOAHOMVADLLEI RONDHRETNELQKSLKEENKGRISA	194	91.28	79.51	16
Finola_ref_p_CaFN_010003730	1	MGKGVLEIR ENI INQVFSKRRNLKAVELHMLDCAEVL I FSSRQLYFEGSSNTI VERRYSTALONNVEGEPNLI LEI PLKAKYVES OTTERPHENIEELGKELQNEKLDRTLSOAHOMVADLLEI RONDHRETNELQKSLKEENKGRISA	244	99.00	100.00	1
LAConfidential_ref_p_CaLAC_000014400	1	MGKGVLEIR ENI INQVFSKRRNLKAVELHMLDCAEVL I FSSRQLYFEGSSNTI VERRYSTALONNVEGEPNLI LEI PLKAKYVES OTTERPHENIEELGKELQNEKLDRTLSOAHOMVADLLEI RONDHRETNELQKSLKEENKGRISA	155	62.30	63.52	36
LAConfidential_ref_p_CaLAC_000015310	1	MGKGVLEIR ENI INQVFSKRRNLKAVELHMLDCAEVL I FSSRQLYFEGSSNTI VERRYSTALONNVEGEPNLI LEI PLKAKYVES OTTERPHENIEELGKELQNEKLDRTLSOAHOMVADLLEI RONDHRETNELQKSLKEENKGRISA	82	62.93	32.38	11
PBBKush_ref_p_CaPBB_000004690	1	MGKGVLEIR ENI INQVFSKRRNLKAVELHMLDCAEVL I FSSRQLYFEGSSNTI VERRYSTALONNVEGEPNLI LEI PLKAKYVES OTTERPHENIEELGKELQNEKLDRTLSOAHOMVADLLEI RONDHRETNELQKSLKEENKGRISA	144	98.61	59.02	2
PurpleKush_ref_p_CaPK_000009050	1	MGKGVLEIR ENI INQVFSKRRNLKAVELHMLDCAEVL I FSSRQLYFEGSSNTI VERRYSTALONNVEGEPNLI LEI PLKAKYVES OTTERPHENIEELGKELQNEKLDRTLSOAHOMVADLLEI RONDHRETNELQKSLKEENKGRISA	225	93.78	90.98	11
PurpleKush_ref_p_CaPK_010001180	1	MGKGVLEIR ENI INQVFSKRRNLKAVELHMLDCAEVL I FSSRQLYFEGSSNTI VERRYSTALONNVEGEPNLI LEI PLKAKYVES OTTERPHENIEELGKELQNEKLDRTLSOAHOMVADLLEI RONDHRETNELQKSLKEENKGRISA	244	100.00	100.00	0

c) OG0000433 - Clade TM3

NCBI Multiple Sequence Alignment Viewer, Version 1.26.0

Sequence ID	Start	Alignment	End	Identity	Coverage	Mismatches
Risteveki_c10_MADS_REF XP_030501771.2	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	237	100.00	100.00	0
LAConfidential_ref_p_CaLAC_000014960	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	181	74.26	76.79	6
Chemdog91_ref_p_CaCD91_000014450	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	87	47.10	34.60	17
CBDx18_ref_p_CaCDB_050001490	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	120	46.51	50.21	59
Cannatonic_ref_p_CaCAN_000014240	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	120	46.51	50.21	59
LAConfidential_ref_p_CaLAC_000008990	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	120	46.51	50.21	59
LAConfidential_ref_p_CaLAC_000016090	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	164	45.04	65.63	56
Cannatonic_ref_p_CaCAN_000014890	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	156	39.08	65.40	87
pink_pepper_ref_p_CaPP_000071452.1	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	160	52.91	67.51	69
CBDx18_ref_p_CaCDB_020003500	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	181	74.26	76.79	6
LAConfidential_ref_p_CaLAC_000005960	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	182	74.26	76.79	6
LAConfidential_ref_p_CaLAC_000010790	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	181	74.26	76.79	6
Finola_ref_p_CaFN_000003390	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	205	43.98	81.86	88
CBDx18_ref_p_CaCDB_020003500	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	221	52.28	88.61	84
PurpleKush_ref_p_CaPK_010001590	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	217	47.48	91.14	103
Risteveki_c10_MADS_REF XP_030502139.1	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	217	47.48	91.14	103
CBDx18_ref_p_CaCDB_050001840	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	217	47.48	91.14	103
pink_pepper_ref_p_CaPP_00002139.1	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	217	47.48	91.14	103
Finola_ref_p_CaFN_010001630	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	217	47.48	91.14	103
LAConfidential_ref_p_CaLAC_000005960	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	228	53.63	91.56	84
PBBKush_ref_p_CaPBB_000004690	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	228	53.63	91.56	84
pink_pepper_ref_p_CaPP_030496578.1	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	228	53.63	91.56	84
LAConfidential_ref_p_CaLAC_000005781	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	228	53.63	91.56	84
Risteveki_c10_MADS_REF XP_030496578.1	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	228	53.63	91.56	84
PurpleKush_ref_p_CaPK_060001800	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	229	53.63	91.56	84
pink_pepper_ref_p_CaPP_06007161.1	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	232	97.89	97.89	0
Finola_ref_p_CaFN_010001910	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	234	89.87	98.73	21
pink_pepper_ref_p_CaPP_06007161.1	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	234	89.87	98.73	21
pink_pepper_ref_p_CaPP_06007161.1	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	235	99.16	99.16	0
pink_pepper_ref_p_CaPP_000001450	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	237	100.00	100.00	0
PurpleKush_ref_p_CaPK_000004560	1	MKRGKTKRI ENAQRSTFSKRRNLKAVELHMLDCAEVL I FSPKRLYFEGSSNTI NITI ERYTORTKDGSGTSKTI SEDDHQLQVAKESFSMAKI DLEVSKMLLQGLSCSFELH ENQILRSAR RANQNYREI EKLKESKSLLENHVAL	237	99.58	100.00	1

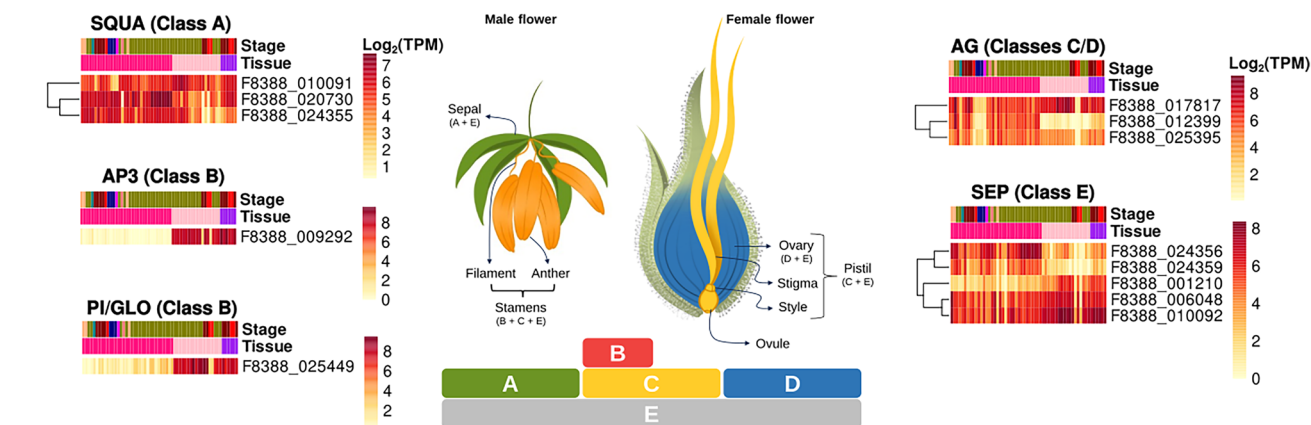
Fig. 2 Multiple sequence alignment showing examples of a) a clade composed entirely of genes classified as MADS-box, and two clades containing b) three non-MADS genes and c) eight non-MADS genes. Non-MADS genes are marked with blue rectangles

of our phylogenomic pipeline—driven by PlantTFDB prediction and OrthoFinder orthology inference—rather than authentic evolutionary divergence. As detailed in Supplementary Table S8, the genomes analyzed in this study differ substantially in assembly contiguity and predicted proteome size, directly shifting the baseline for gene discovery. High-quality assemblies such as Pink Pepper (scaffold N50: 77 Mb; 39,975 proteins) and CBDx18 (scaffold N50: 91.9 Mb; 33,371 proteins) contrast sharply with more fragmented assemblies such as Chemdog91 (scaffold N50: 2.3 kb; 16,538 proteins) and LA Confidential (scaffold N50: 2.6 kb; 24,494 proteins), despite the latter having high sequencing coverage (300x and 50x, respectively). Importantly, the “chromosome-level” classification can be misleading for Purple Kush (scaffold N50: 133.9 kb; 12,836 contigs) and Finola (scaffold N50: 370.5 kb; 5,303 contigs), which, despite their assembly-level designation, exhibit contiguity metrics more typical of contig- or scaffold-level assemblies. In highly fragmented assemblies, gene predictions within specific orthogroups are frequently truncated or missed, particularly for gene families prone to tandem duplication such as SEP, thereby reducing the number of genes recoverable by

homology-based and phylogenetic approaches. The apparent absence of conserved clades such as AP3 in some cultivars is therefore more likely an artifact of annotation incompleteness than a true biological absence. Therefore, resolving these genomic discrepancies is important to establish a reliable and stable basis for the ABCDE model of Cannabis. By demonstrating that these central regulators of floral identity are evolutionarily conserved across varieties, despite prominent assembly artifacts, we provide the community with a reliable and validated catalog of target genes. This report can prevent future functional studies and molecular breeding programs from misinterpreting technical annotation gaps as genuine losses of lineage-specific genes, ensuring that efforts to optimize sexual expression and agronomic traits of the species are anchored in a robust evolutionary orthology.

We identified 60 MADS-box TFs in CannAtlas, representing all clades described by Risteveki (2023), except for AGL15 (Figure 3). Genes from the AGL6, FLC-like, and Bsister clades, as well as SEP genes F8388_024356 and F8388_024359, and the AG gene F8388_012399, showed high expression in FF (Figure 3). In contrast, all genes from the AP3, PI/GLO, SVP, and MIKCS clades, along

a) Cannabis ABCDE model scheme and expression of ABCDE MADS-Box genes



b) Expression of other MADS-Box genes

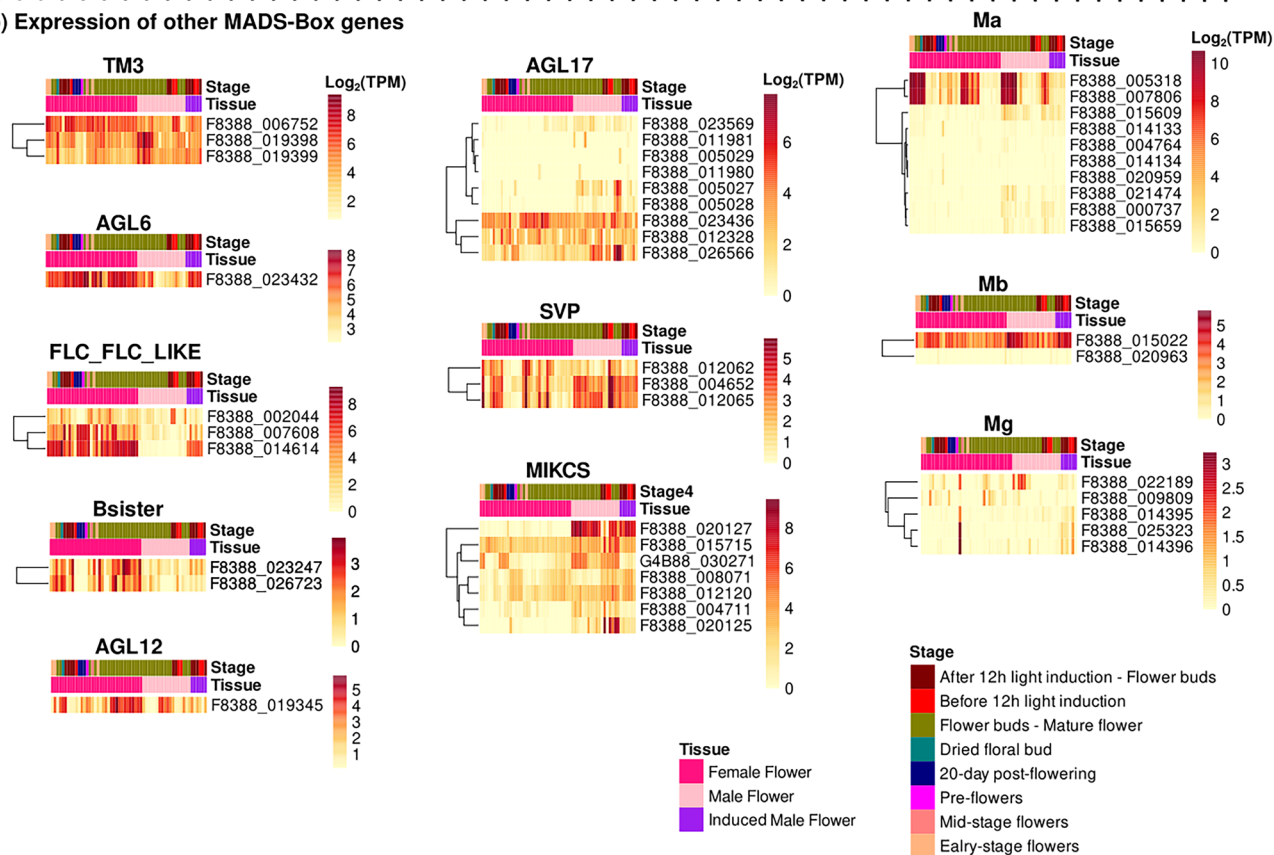


Fig. 3 Expression profiles of representative MADS-box genes by clade. Heatmap shows normalized expression for representative genes from each orthology-defined clade. The “Class” column refers to the ABCDE model of floral organ identity, as represented at the central scheme. The “Stage” column indicates the developmental stage for

each sample, as reported in the SRA metadata. The central scheme depicts the ABCDE model; colors map classes to organs. Model adapted from Y.-T. Chen et al., (2019); flower drawings adapted from Releaf (2023), Schilling et al. (2021) and Yadav et al. (2023). Illustration: Rosana Gobbi Vettorazzi.

with the SEP gene F8388_001210, are predominantly expressed in MF (Figure 3). The remaining AG and SEP genes, as well as representatives from other MADS-box clades, exhibit relatively balanced expression between MF and FF (Figure 3). The differential expression of SEP

homologues between male and female flowers is also reported by Garcia-de Heer et al. (2025), who identified six SEP homologues in cannabis with distinct sex-biased expression patterns and proposed potential subfunctionalization of the SEP subfamily in sex determination.

Experimental support for a role of SEP genes in *C. sativa* floral development comes from Adal et al. (2025), who showed that overexpression of csSEP1 accelerates flowering in *Arabidopsis thaliana*, suggesting that at least some cannabis SEP genes retain ancestral regulatory functions, even as others may have undergone subfunctionalization toward sex-specific roles as proposed by Garcia-de Heer et al. (2025). Our orthology-based analysis across nine *C. sativa* genomes complements these findings by placing the differentially expressed SEP genes within conserved orthogroups, offering a comparative genomic perspective on their evolutionary relationships and functional diversification.

According to the ABCDE model of floral organ identity (F. Chen et al. 2017; Dreni & Kater 2014; H. Liu et al. 2021a, b), the major classes of MADS-box genes are defined as follows:

- Class A: SQUAMOSA (SQUA)/APETALA1 (AP1);
- Class B: APETALA3 (AP3) and PISTILLATA (PI/GLO);
- Class C: AGAMOUS (AG);
- Class D: SHATTERPROOF (SHP) and SEEDSTICK (STK), (from AG subfamily);
- Class E: SEPALLATA (SEP).

In addition to these core classes, other MADS-box genes such as AGL6, FLC, FLC-like, AGL17, TM3, SVP, MIKCS, AGL12, Bsister, and AGL15 are also implicated in flowering processes (H. Liu et al. 2021a, b; Ristevski 2023; Shah et al. 2022).

The ABCDE model suggests that the male reproductive organ—the stamen (anther and filament)—is specified by the combined action of class B, C, and E genes (Figure 3) (Gioppato & Dornelas 2019). Based on the expression patterns observed here (Figure 3), we propose that the class B genes AP3 (F8388_009292) and PI/GLO (F8388_025449), the class C/D AG (F8388_017817), and class E SEP (F8388_001210) are strong candidates for regulators of MF identity in *C. sativa*. This is consistent with Garcia-de Heer et al. (2025), who reported preferential expression of AP3 and PI homologues in male flowers and their downregulation following ethephon-induced feminization, suggesting a direct role of class B genes in specifying unisexual floral identity in cannabis. While their study focused on a single chemotype under ethephon treatment, our analysis integrates 117 RNA-Seq samples across multiple chemotypes and genotypes, providing a broader transcriptomic context that corroborates and extends these findings. Additionally, the MIKCS genes—particularly F8388_020127—also show higher expression in MF, suggesting a possible role in MF development (Figure 3). Although MIKCS genes have been previously linked to pollen development (Gramzow &

Theissen 2010), their roles in broader aspects of floral morphogenesis remain poorly understood.

MADS-box genes from classes C, D, E, and Bsister are known to be involved in the development of female reproductive organs such as the ovary, style, and stigma (Figure 3) (de Folter et al. 2006; Gioppato & Dornelas 2019). Our data support that the C/D gene AG (F8388_012399), the E genes SEP (F8388_024356 and F8388_024359), and Bsister genes (F8388_023247 and F8388_026723) are likely regulators of FF identity in cannabis (Figure 3). In addition, AGL6 (F8388_023432) and FLC-like (F8388_007608 and F8388_014614) genes also exhibit higher expression in FF, further suggesting their involvement in FF development (Figure 3).

Although experimental validation is required to confirm the functional roles of these candidate genes, our integrative analysis using publicly available transcriptomic data provides a valuable foundation for prioritizing MADS-box TFs involved in sex-specific floral development in *C. sativa*.

Identification and expression patterns of candidate genes involved in pollen development

To investigate genes beyond the canonical MADS-box regulators involved in MF development in *C. sativa*, we conducted BLAST searches against CannAtlas using three reference gene sets. First, from the 62 sugar metabolism-related genes reported in maize by Liu et al. (2021a, b), we identified 35 homologs in CannAtlas, of which 8 displayed preferential expression in both genetic MF and IMF (Figure 4, Table S4). Second, among the 61 candidate genes proposed by Adal et al. (2021) as associated with masculinization in cannabis, 37 homologs were detected in CannAtlas, with 9 showing preferential expression in MF and IMF (Figure 4, Table S4). Third, from the 185 reference genes implicated in male sterility in Chinese cabbage by Huang et al. (2020), we identified 42 homologs in CannAtlas, of which 19 exhibited preferential expression in MF and IMF (Figure 4, Table S4).

Huang et al. (2020) grouped their gene set into four functional categories: pollen development and tube growth, pollen wall development, phytohormone regulation, and TFs. Similarly, Adal et al. (2021) classified their candidates by specific functions in flower development. Building on these frameworks, we assigned the 35 CannAtlas genes identified in our analysis to 9 functional categories (FC) related to flower and pollen development (Figure 4, Table S5).

Notably, nearly all genes under “pollen development and pollen tube growth”, together with approximately two-thirds of those in “pollen wall development”, exhibit high expression in MF samples collected after 12 hours of light induction—coinciding with the onset of floral development. A similar expression pattern was also observed in

Expression of pollen development genes

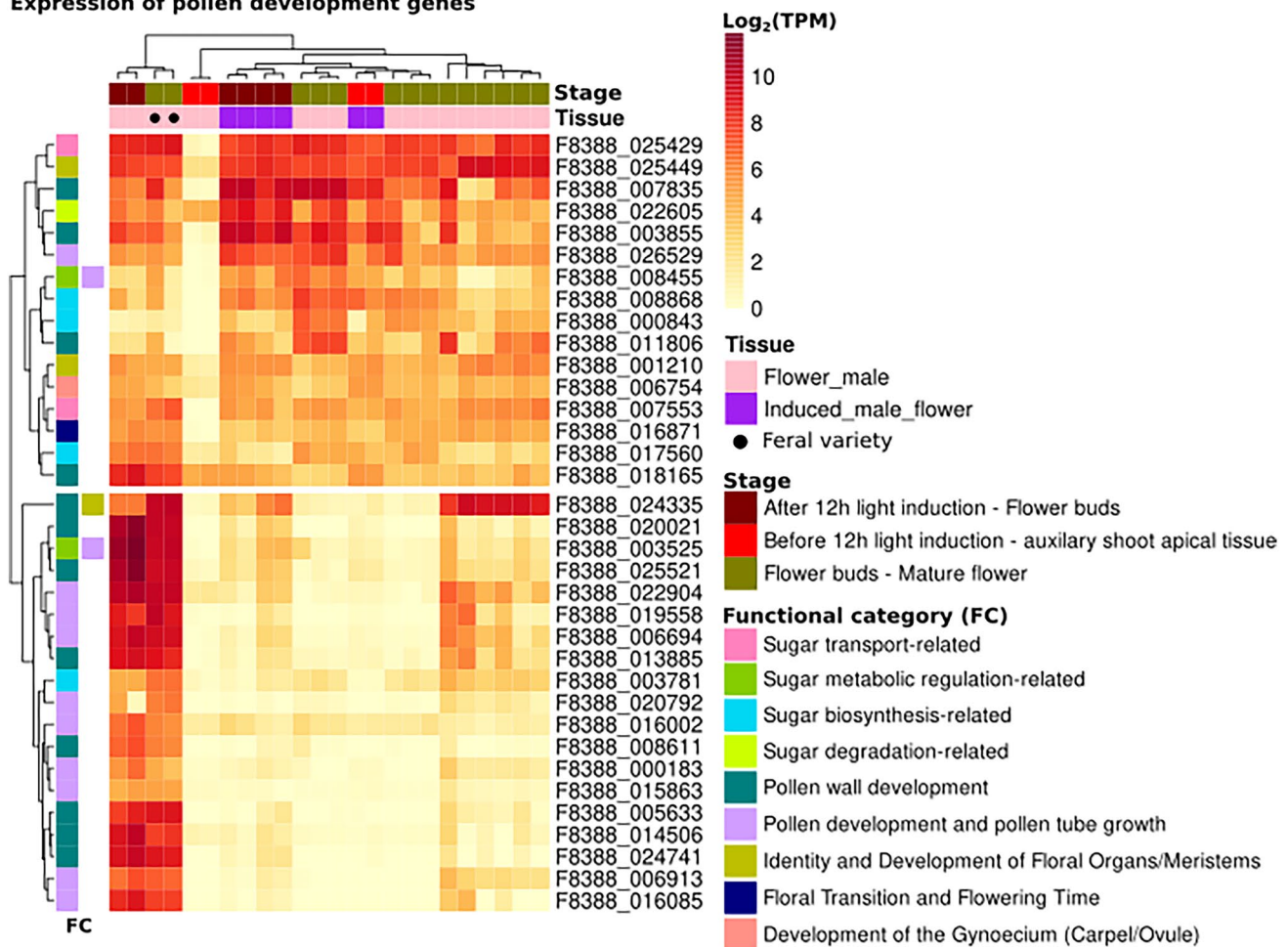


Fig. 4 Expression patterns of genes associated with pollen development, classified into functional categories (FCs). Only genes showing preferential expression in male flowers (MF) and/or induced male flowers (IMF) relative to other tissues in the Cannabis Express-

ion Atlas (CannAtlas) are shown (see Supplementary Figure 18 for expression profiles across all CannAtlas tissues). “Stage” refers to the developmental stage of the samples as reported in the SRA database

MF samples from two feral varieties (SAMN20863695 and SAMN20863694). In contrast, 16 genes across multiple categories show comparable expression levels in both MF and IMF, with most showing little to no expression in MF samples collected before the 12 h light induction mark (i.e., at the axillary shoot apical meristem stage). This expression pattern suggests the existence of a light-triggered developmental switch that initiates a coordinated gene expression program essential for MF development, including genes involved in pollen formation.

Among the 16 genes expressed similarly in MF and IMF were four MADS-box TFs. Two of these—F8388_001210 (class E, SEP) and F8388_025449 (class B, PI/GLO)—are canonical ABCDE model genes that govern floral organ identity and are essential for stamen specification. The other two—F8388_006754 and F8388_016871—are non-ABCDE MADS-box genes. F8388_006754, a putative AGL6-like

gene (58.98% similarity to OsMADS6 (Table S4)), has been associated with gynoecium development (Table S5), while F8388_016871, a putative AGL15-like gene (56.45% similarity to AGL15_ARATH (Table S4)), has been implicated in floral transition and regulation of flowering time (Table S5). Interestingly, although AGL6-like (F8388_006754) is typically linked to female organ formation, its expression in MF suggests it may play a repressive role in this context. A similar mechanism has been described in *Silene latifolia*, where Hardenack et al. (1994) and Bačovský et al. (2022) showed that altered expression of MADS-box genes, including class B genes PISTILATA and APETALA3, correlates with repression of gynoecium development in MF. Likewise, Di Stilio et al. (2005) proposed that in dioecious species, sex determination may involve modulation of floral homeotic gene activity, particularly through the downregulation of female organ identity programs during male flower

development. Taken together, these findings support the hypothesis that AGL6-like (F8388_006754) contributes to repressing gynoecium formation in male *C. sativa* flowers, thereby facilitating proper stamen development.

Among the 9 sugar-related genes identified, seven displayed similar expression levels in both MF and IMF, suggesting a conserved role in pollen development across genetic and chemically induced contexts. Two of these genes encode sugar transporters from the SWEET family (F8388_007553 and F8388_025429). Along this, other two transporters classified with mixed expression at CanAtlas also had considerable expression in male flowers: a sucrose/H⁺ symporter (F8388_000867), and a UDP-galactose/UDP-glucose transporter (F8388_001017) (Table S4). These transporters likely help sustain the high-energy demands for pollen development and pollen tube elongation. In *Arabidopsis*, disruption of sucrose transport impairs pollen germination and shortens pollen tubes, underscoring the critical role of sugar supply in reproductive success (De Angeli 2022; Reinders 2016). In addition, two genes associated with sugar metabolism regulation—a putative L-ascorbate oxidase (F8388_003525) and a MYB TF (F8388_008455)—may contribute to modulating redox balance and secondary metabolism during pollen development, processes that influence osmotic regulation and energy availability, both essential for pollen tube growth (Li et al. 2020). We also identified two hexosyltransferases (F8388_000843 and F8388_008868) and a 1,3- β -glucan synthase (F8388_017560) involved in sugar biosynthesis.

Hexosyltransferases catalyze the transfer of hexoses to elongating polysaccharide chains, contributing to the formation of cell wall components such as hemicelluloses and pectins (Li et al. 2020). Meanwhile, 1,3- β -glucan synthases drive callose production, a structural component of the pollen wall that reinforces integrity during maturation (Nishikawa et al. 2005). Collectively, these genes likely coordinate dynamic remodeling of the pollen cell wall through the regulated synthesis and deposition of carbohydrate polymers, ensuring structural stability and functionality in both MF and IMF.

Five genes related to pollen wall development also showed comparable expression levels in both MF and IMF, supporting a shared molecular program for pollen wall formation across both flower types. These include a cytochrome P450 (F8388_003855), a NAD-dependent epimerase/dehydratase (F8388_007835), a chalcone synthase (F8388_011806), a beta-galactosidase (F8388_024335), and a pectinesterase (F8388_018165). These enzymes act in the biosynthesis and remodeling of the structurally complex pollen wall layers. Cytochrome P450 enzymes contribute to anther cuticle development and sporopollenin biosynthesis, both essential for the outer exine layer (Ma et al. 2022; Yang et al. 2014). NAD-dependent epimerases/dehydratases facilitate sugar residue modification for polysaccharide biosynthesis (Usadel et al. 2004). Chalcone synthase catalyzes the first committed step in flavonoid biosynthesis, critical for pollen viability and wall integrity (Buer et al. 2010). Beta-galactosidases participate in cell wall remodeling by cleaving galactose-containing polysaccharides, enabling proper wall expansion

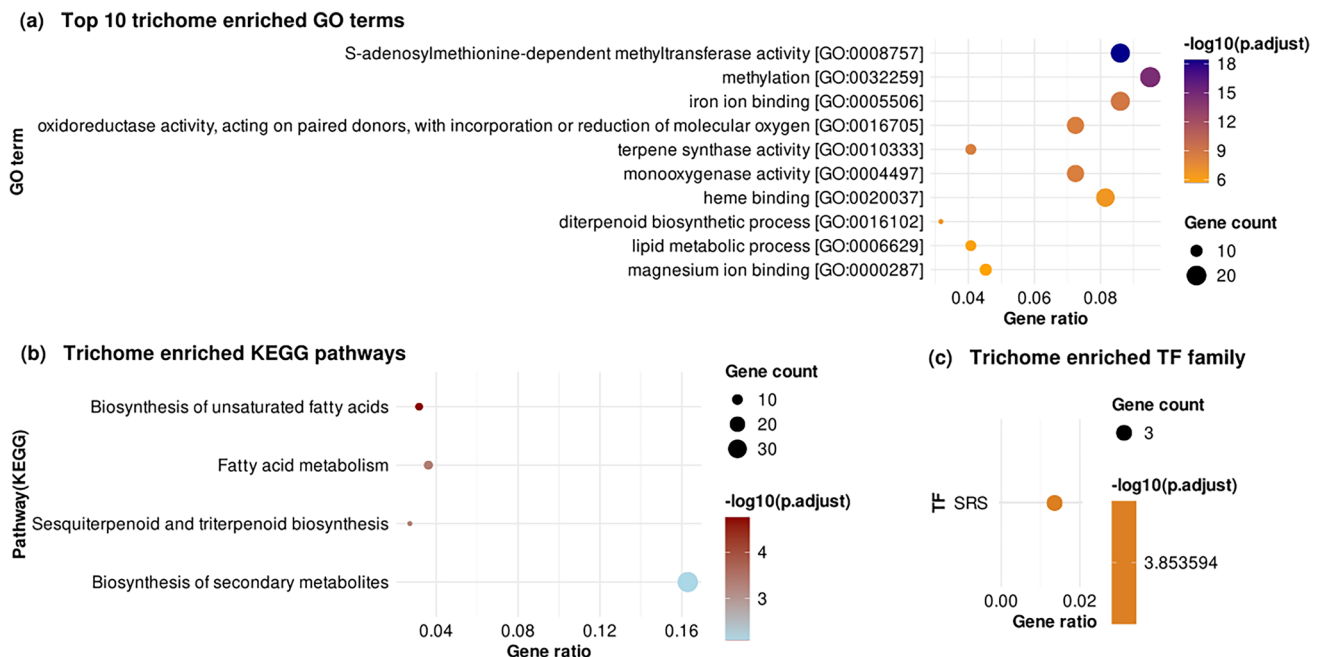


Fig. 5 Functional enrichment of trichome-specific genes. **a** Gene ontology (GO) terms, **b** KEGG metabolic pathways and **(c)** transcription factor (TF) families

and maturation (Ban et al. 2020). Pectinesterases modulate pectin methylesterification, thereby adjusting wall porosity, rigidity, and elasticity (Cankar et al. 2014). Additionally, some genes appear to contribute to both pollen development and pollen tube growth, such as a short-chain dehydrogenase/reductase (SDR, F8388_026529) and the previously mentioned MYB TF (F8388_008455), highlight the coordinated regulation of wall biosynthesis and signaling pathways required for male gametophyte function.

Upstream cannabinoid and terpenoid pathway genes show parallel expression in hemp and marijuana

The FF of *C. sativa* have historically been the most utilized part of the plant due to their psychotropic and medicinal properties. In plant physiology, secondary metabolite production is commonly associated with defense responses against biotic and abiotic stresses. In cannabis, this defensive role is largely mediated by glandular trichomes—specialized epidermal structures on FF that synthesize and store a wide array of secondary metabolites, including cannabinoids and terpenoids.

To gain deeper insight into the functional roles of trichome-specific genes identified in CannAtlas (Barbosa-Xavier et al. 2024), we performed enrichment analyses for Gene Ontology (GO), metabolic pathway (KEGG), and TF families. The enriched GO and KEGG terms strongly point to the biosynthesis, modification, and storage of secondary metabolites (Figure 5).

Among the enriched GO terms (Figure 5a), S-adenosylmethionine-dependent methyltransferase activity and methylation are particularly relevant, reflecting the role of methylation in enhancing the bioactivity, chemical stability, and defense potential of secondary metabolites such as cannabinoids and terpenoids (C. Zhang et al. 2021). Additional enrichment in ion binding, heme binding, and monooxygenase activity (Figure 5a) highlights the contribution of cytochrome P450 enzymes, which participate in the structural diversification of secondary metabolites and modulate their chemical properties (Layer et al. 2010). The enrichment of oxidoreductase activity (Figure 5a) underscores the oxidative processes necessary for synthesizing bioactive compounds. Terpene synthase activity (Figure 5a) was also significantly enriched, consistent with the role of volatile terpenes in repelling herbivores and inhibiting microbial pathogens (MacWilliams et al. 2023). GO terms related to the diterpenoid biosynthesis and lipid metabolism (Figure 5a) support the production of cannabinoids and their fatty acid precursors.

KEGG enrichment analysis (Figure 5b) corroborated these patterns. Enrichment of unsaturated fatty acid biosynthesis and fatty acid metabolism point to the generation

of precursors for signaling molecules like jasmonic acid, a key hormone in plant defense (Luo et al. 2019; Marks et al. 2009). The enrichment of sesquiterpenoid and triterpenoid biosynthesis highlights the trichome role in producing antimicrobial and insect-repellent compounds, while the broad enrichment of secondary metabolite biosynthesis emphasizes the specialized metabolic capacity of trichomes.

TF enrichment (Figure 5c) revealed the SRS (SHI-related sequence) family as significantly overrepresented. This TF family regulates both trichome development and secondary metabolism (J. Zhang et al. 2020), suggesting that SRS TFs coordinate structural and chemical components of the cannabis defense system. Taken together, these results highlight cannabis glandular trichomes as hubs of defensive metabolism, integrating gene regulation, specialized metabolite biosynthesis, and adaptive responses to environmental challenges.

Cannabinoids and terpenes are the most extensively investigated compounds in *C. sativa* for their medicinal or recreational importance. A detailed understanding of their metabolic pathways is essential for strategies aimed at manipulating metabolite content in cannabis plants. To identify key genes involved in these pathways, we performed BLAST searches in CannAtlas using proteins previously characterized in cannabinoid and terpenoid biosynthesis (Kovalchuk et al. 2020), including those from upstream pathways such as polyketide, methylerythritol (MEP), and mevalonate (MEV) (Kovalchuk et al. 2020). This search identified 42 genes (Table S6). Given that these secondary metabolites are synthesized and stored in glandular trichomes of FF, we assessed their expression in flower and trichome samples (Figure 6).

The polyketide pathway, which produces olivetolic acid, involves three core enzymes: Acyl-Activating Enzymes (AAEs), Olivetol Synthase (OLS), and Olivetolic Acid Cyclase (OAC), with OLS and OAC being directly implicated in cannabinoid biosynthesis (Thomas et al. 2020). We identified 13 polyketide-related genes (Table S6), including six AAEs classified as “expressed-in-all” and two as “mixed” in CannAtlas. Two OLS genes (F8388_010207, F8388_010208) and two OAC genes (F8388_012844, F8388_012845) showed trichome-specific expression, while another OLS gene (F8388_010206) was enriched in trichomes and IMF (Table S7). Of the eight MEP pathway genes identified (Table S6), four were expressed in all samples, two showed trichome-specific expression—including HDS (F8388_013291) and HDR (F8388_013980)—and one was leaf-specific (Table S7). All six MEV pathway genes were classified as “expressed-in-all”. Notably, isopentenyl-diphosphate delta-isomerase (IDI), which interconverts the MEP- and MEV-derived IPP and DMAPP to form the isoprenoid backbone (Krause et al. 2023), was represented by two genes, one of which (F8388_016754) was

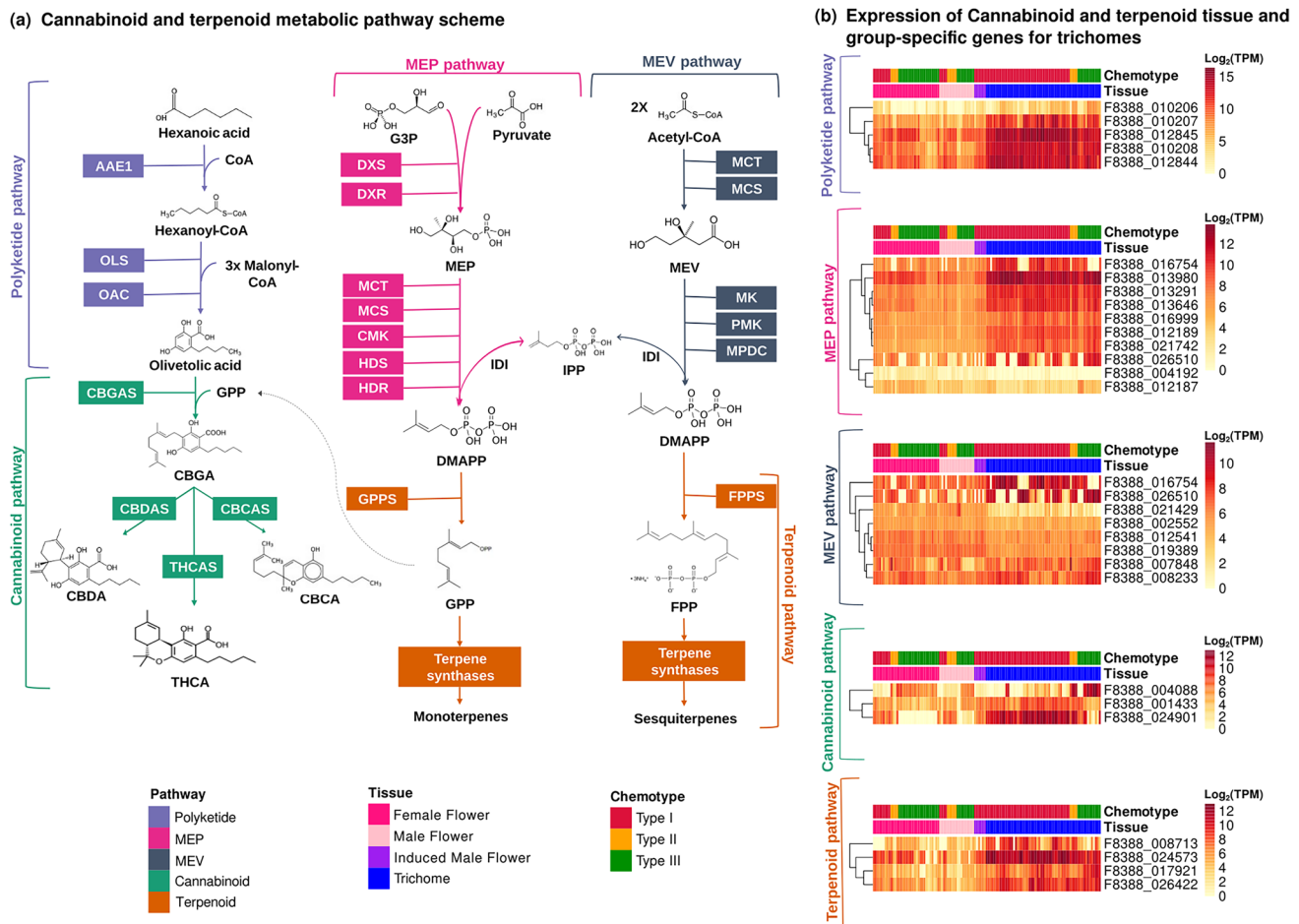


Fig. 6 Cannabinoid and terpenoid biosynthesis. **a** Schematic representation of the metabolic pathways underlying cannabinoid and terpenoid production, with key enzymatic steps highlighted. **b** Expression profiles of cannabinoid- and terpenoid-related genes in trichome

and whole-flower samples. Expression levels are shown across chemotypes to illustrate tissue specificity and chemotype-associated patterns.

trichome-specific. Together, these results indicate that genes from the polyketide and MEP pathways are preferentially expressed in trichomes. Importantly, no consistent expression differences were observed between hemp and marijuana varieties, nor between MF and FF, suggesting conservation of upstream cannabinoid biosynthetic pathways across chemotypes and sexual morphs.

We identified five genes directly involved in the cannabinoid biosynthetic pathway: three prenyltransferases (F8388_001433, F8388_008405, and F8388_013365), one Cannabidiolic Acid Synthase (CBDAS, F8388_004088), and one Tetrahydrocannabinolic Acid Synthase (THCAS, F8388_024901) (Table S6). Of these, THCAS and the prenyltransferase F8388_001433 were trichome-specific; CBDAS is group-enriched in FF and stems; the prenyltransferase F8388_008405 is mixed; and F8388_013365 is expressed-in-all tissues (Table S7). The classification of CBDAS as group-enriched in FF and stems, rather than

trichome-specific, likely reflects both chemotype distribution and the methodology used by Barbosa-Xavier et al. (2024), which relies on median expression across samples. Approximately 70% of the trichome samples in the dataset originated from type I chemotypes, which typically show low CBDAS expression, potentially underestimating its trichome specificity. A similar bias could have occurred for THCAS had the dataset been skewed toward type III chemotypes. These observations highlight how sample composition and classification thresholds can affect the apparent tissue specificity of cannabinoid biosynthesis genes.

Cannabigerolic acid (CBGA), the central precursor of cannabinoids, is synthesized via alkylation of OAC and GPP, a reaction catalyzed by a CsPT1 aromatic prenyltransferase (CBGAS) (Innes & Vergara 2023). The gene F8388_001433 shows high sequence identity to several prenyltransferases (Table S6), including CsPT7 (98.4%), CsPT1 (83.9%), and CsPT4 (78.7%). CsPT4 and CsPT7 are closely related to

CsPT1, and CsPT4 has also been shown to catalyze CBGA formation in vivo (Luo et al. 2019; Rea et al. 2019). These similarities strongly suggest that F8388_001433 functions as an aromatic prenyltransferase involved in CBGA biosynthesis. Expression data support this functional assignment: THCAS, CBDAS, and CBGAS (F8388_001433) were more highly expressed in trichomes than in whole flowers (Figure 6b). Moreover, THCAS and CBGAS were preferentially expressed in type I and II chemotypes, while CBDAS was more highly expressed in type II and III chemotypes (Figure 6b), consistent with their roles in defining cannabinoid profiles.

Unexpected expression patterns were detected in certain type III hemp samples. Trichome samples SAMN09747685, SAMN09747688, and SAMN09747691, labeled as type III hemp (Finola strain) (Livingston et al. 2020), displayed type II-like expression profiles (Figure 6b). These samples are biological replicates from BioProject PRJNA483805, which otherwise shows consistent type III expression. We hypothesize that these samples may have been inadvertently contaminated with trichomes from type I varieties, such as Purple Kush or Hindu Kush, which were analyzed alongside Finola by Livingston et al. (2020) using two-photon microscopy and disc cell counting analysis. A second discrepancy was observed in the samples SAMN42101035, SAMN42101034, and SAMN42101033 from the MW6-15 cultivar (BioProject PRJNA1128734). Although described by Welling et al. (2023) as an industrial hemp line (type III), these samples displayed a type I expression profile, characterized by high THCAS expression and complete absence of CBDAS (Figure 6b). Such inconsistencies may reflect sample mislabeling, cross-contamination, or previously unrecognized chemotypic diversity, and warrant further validation.

Regarding terpenoid biosynthesis, we identified eight candidate genes (Table S6): three with root-specific expression, three with trichome-specific, one group-enriched in trichomes and IMF, and one with low expression (Table S7). Trichome-specific terpene synthase genes (F8388_008713, F8388_024573, F8388_026422) and the group-enriched gene (F8388_017921) all exhibited markedly higher expression in trichomes compared to whole-flower samples (Figure 6). Importantly, their expression showed no clear bias toward specific chemotypes. This pattern suggests that terpenoid biosynthesis is broadly active across both hemp and marijuana varieties, likely independent of cannabinoid chemotype-defining genes.

Overall, despite chemotypic differences between hemp and marijuana, genes in the upstream cannabinoid and terpenoid pathways show similar expression profiles (Figure 6), highlighting conserved mechanisms of defense specialization in *C. sativa*.

Conclusion

This study advances our understanding of *C. sativa* flower development by identifying OGs of MADS-box genes and analyzing their expression in MF and FF. Our results point to the involvement of specific MADS-box genes in flower organ development and sex differentiation. In parallel, sugar metabolism-related genes were linked to fertility and pollen wall formation, underscoring key metabolic processes in male reproductive structures. Expression profiling of cannabinoid and terpenoid biosynthetic genes further confirmed their strong tissue specificity, particularly in trichomes, emphasizing the finely tuned regulation of secondary metabolite production. Collectively, these findings establish a molecular framework for future functional studies and identify promising targets for metabolic engineering and breeding strategies aimed at optimizing flower traits and bioactive compound profiles in cannabis.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00425-026-05043-6>.

Acknowledgements This work was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES; Finance Code 001), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Programa de Apoio à Pesquisa, Inovação e Cultura (PAPIC - UENF). The funding agencies had no role in the design of the study and collection, analysis and interpretation of data and in writing.

Author contributions Conceived the study: K.B.-X. and T.M.V.; Funding and resources: T.M.V.; Data analysis: K.B.-X.; Interpretation of the results: K.B.-X. and T.M.V.; Wrote the manuscript: K.B.-X. and T.M.V.

Funding The Article Processing Charge (APC) for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614). Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES; Finance Code 001); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Universidade Estadual do Norte Fluminense Darcy Ribeiro.

Data availability Not applicable.

Declarations

Conflict of interest Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adal AM, Doshi K, Holbrook L, Mahmoud SS (2021) Comparative RNA-Seq analysis reveals genes associated with masculinization in female *Cannabis sativa*. *Planta*. <https://doi.org/10.1007/S00425-020-03522-Y>
- Adal AM, Sajaditabar R, Nguyen T, Mahmoud SS (2025) Differential gene expression analysis reveals transcription factors that control flower initiation and development in *Cannabis sativa*. *Biocatal Agric Biotechnol* 67:103691. <https://doi.org/10.1016/j.bcab.2025.103691>
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- De Angeli A (2022) Pollen likes sugars: sucrose-specific transport by AtSWEET13. *Proc Natl Acad Sci U S A* 119(46):e2216610119. <https://doi.org/10.1073/pnas.2216610119>
- Bačovský V, Čegan R, Tihlaříková E, Neděla V, Hudzieczek V, Smrža L, Janíček T, Beneš V, Hobza R (2022) Chemical genetics in *Silene latifolia* elucidate regulatory pathways involved in gynoecium development. *J Exp Bot* 73(8):2354–2368. <https://doi.org/10.1093/jxb/erab538>
- van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE (2011) The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol* 12(10):R102. <https://doi.org/10.1186/gb-2011-12-10-r102>
- Ban Q, Jiao J, He Y, Jin M, Rao J (2020) Ectopic expression of the persimmon β -galactosidase gene *DkGAL2* promotes leaf growth, delays dark-induced senescence and enhances tolerance to abiotic stress in *Arabidopsis*. *Sci Hortic* 265:109232. <https://doi.org/10.1016/j.scienta.2020.109232>
- Barbosa-Xavier K, Pedrosa-Silva F, Almeida-Silva F, Venancio TM (2024) Cannabis expression atlas: a comprehensive resource for integrative analysis of *Cannabis sativa* L. gene expression. *Physiol Plant* 176(6):e70010. <https://doi.org/10.1111/ppl.70010>
- Buer CS, Imin N, Djordjevic MA (2010) Flavonoids: new roles for old molecules. *J Integr Plant Biol* 52(1):98–111. <https://doi.org/10.1111/j.1744-7909.2010.00905.x>
- Cankar K, Kortstee A, Toonen MAJ, Wolters-Arts M, Houben R, Mariani C, Ulvskov P, Jorgensen B, Schols HA, Visser RGF, Trindade LM (2014) Pectic arabinan side chains are essential for pollen cell wall integrity during pollen development. *Plant Biotechnol J* 12(4):492–502. <https://doi.org/10.1111/pbi.12156>
- Chen F, Zhang X, Liu X, Zhang L (2017) Evolutionary Analysis of MIKCC-Type MADS-Box Genes in Gymnosperms and Angiosperms. *Front Plant Sci* 8:895. <https://doi.org/10.3389/fpls.2017.00895>
- Chen Y-T, Chang C-C, Chen C-W, Chen K-C, Chu Y-W (2019) MADS-Box Gene Classification in Angiosperms by Clustering and Machine Learning Approaches. *Front Genet* 9:707. <https://doi.org/10.3389/fgene.2018.00707>
- Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* 353(6339):31–37. <https://doi.org/10.1038/353031a0>
- Dreni L, Kater MM (2014) MADS reloaded: evolution of the AGAMOUS subfamily genes. *New Phytol* 201(3):717–732. <https://doi.org/10.1111/nph.12555>
- ElSohly MA, Slade D (2005) Chemical constituents of marijuana: the complex mixture of natural cannabinoids. *Life Sci* 78(5):539–548. <https://doi.org/10.1016/j.lfs.2005.09.011>
- Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20(1):238. <https://doi.org/10.1186/s13059-019-1832-y>
- Ferber SG, Namdar D, Hen-Shoval D, Eger G, Koltai H, Shoval G, Shbiro L, Weller A (2020) The “entourage effect”: terpenes coupled with cannabinoids for the treatment of mood disorders and anxiety disorders. *Curr Neuropharmacol* 18(2):87–96. <https://doi.org/10.2174/1570159X17666190903103923>
- de Folter S, Shchennikova AV, Franken J, Busscher M, Baskar R, Grossniklaus U, Angenent GC, Immink RGH (2006) A Bister MADS-box gene involved in ovule and seed development in *Petunia* and *Arabidopsis*. *Plant J* 47(6):934–946. <https://doi.org/10.1111/j.1365-313X.2006.02846.x>
- Garcia-de Heer L, Guo Q, Mieog J, Nolan M, Liu L, Dimopoulos N, Melzer R, Kretschmar T (2025) Transcriptomic analysis of ethephon-induced sex reversion of male *Cannabis sativa* reveals changes in expression of floral homeotic genes and a distinct trichome morphology. *J Exp Bot* 76(18):5401–5418. <https://doi.org/10.1093/jxb/eraf291>
- Gioppato HA, Dornelas MC (2019) When Bs are better than As: the relationship between B-class MADS-box gene duplications and the diversification of perianth morphology. *Trop Plant Biol* 12(1):1–11. <https://doi.org/10.1007/s12042-018-9212-4>
- Gramzow L, Theissen G (2010) A hitchhiker’s guide to the MADS world of plants. *Genome Biol* 11(6):214. <https://doi.org/10.1186/gb-2010-11-6-214>
- Grassa CJ, Wenger JP, Dabney C, Poplawski SG, Motley ST, Michael TP, Schwartz CJ, Weiblen GD (2018) A complete Cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. *bioRxiv*. <https://doi.org/10.1101/458083>
- Grassa CJ, Weiblen GD, Wenger JP, Dabney C, Poplawski SG, Timothy Motley S, Michael TP, Schwartz CJ (2021) A new cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytol* 230(4):1665–1679. <https://doi.org/10.1111/nph.17243>
- Gülck T, Möller BL (2020) Phytocannabinoids: origins and biosynthesis. *Trends Plant Sci* 25(10):985–1004. <https://doi.org/10.1016/j.tplants.2020.05.005>
- Hanuš LO, Hod Y (2020) Terpenes/terpenoids in cannabis: are they important? *Med Cannabis Cannabinoids* 3(1):25–60. <https://doi.org/10.1159/000509733>
- Huang S, Peng S, Liu Z, Li C, Tan C, Yao R, Li D, Li X, Hou L, Feng H (2020) Investigation of the genes associated with a male sterility mutant (msm) in Chinese cabbage (*Brassica campestris* ssp. *Pekinensis*) using RNA-Seq. *Mol Genet Genomics* 295(1):233–249. <https://doi.org/10.1007/s00438-019-01618-z>
- Innes PA, Vergara D (2023) Genomic description of critical cannabinoid biosynthesis genes. *Botany* 101(7):270–283. <https://doi.org/10.1139/cjb-2022-0140>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772–780. <https://doi.org/10.1093/molbev/mst010>
- Kovalchuk I, Pellino M, Rigault P, van Velzen R, Ebersbach J, Ashnest JR, Mau M, Schranz ME, Alcorn J, Laprairie RB, McKay JK, Burbridge C, Schneider D, Vergara D, Kane NC, Sharbel TF (2020) The genomics of cannabis and its close relatives. *Annu Rev Plant Biol* 71(1):713–739. <https://doi.org/10.1146/annurev-arpla-081519-040203>
- Krause T, Wiesinger P, González-Cabanelas D, Lackus N, Köllner TG, Klüpfel T, Williams J, Rohwer J, Gershenzon J, Schmidt A (2023) HDR, the last enzyme in the MEP pathway, differently

- regulates isoprenoid biosynthesis in two woody plants. *Plant Physiol* 192(2):767–788. <https://doi.org/10.1093/plphys/kiad110>
- Lavery KU, Stout JM, Sullivan MJ, Shah H, Gill N, Holbrook L, Deikus G, Sebra R, Hughes TR, Page JE, van Bakel H (2019) A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Res* 29(1):146–156. <https://doi.org/10.1101/gr.242594.118>
- Layer G, Reichelt J, Jahn D, Heinz DW (2010) Structure and function of enzymes in heme biosynthesis. *Protein Sci* 19(6):1137–1161. <https://doi.org/10.1002/pro.405>
- Leinen ZJ, Mohan R, Premadasa LS, Acharya A, Mohan M, Byrareddy SN (2023) Therapeutic potential of cannabis: a comprehensive review of current and future applications. *Biomedicine* 11(10):2630. <https://doi.org/10.3390/biomedicine11102630>
- Li C, Meng D, Piñeros MA, Mao Y, Dandekar AM, Cheng L (2020) A sugar transporter takes up both hexose and sucrose for sorbitol-modulated in vitro pollen tube growth in apple[OPEN]. *Plant Cell* 32(2):449–469. <https://doi.org/10.1105/tpc.19.00638>
- Liu H, Yang L, Tu Z, Zhu S, Zhang C, Li H (2021a) Genome-wide identification of MIKC-type genes related to stamen and gynoecium development in *Liriodendron*. *Sci Rep* 11(1):6585. <https://doi.org/10.1038/s41598-021-85927-7>
- Liu S, Li Z, Wu S, Wan X (2021b) The essential roles of sugar metabolism for pollen development and male fertility in plants. *Crop J* 9(6):1223–1236. <https://doi.org/10.1016/j.cj.2021.08.003>
- Livingston SJ, Quilichini TD, Booth JK, Wong DCJ, Rensing KH, Laflamme-Yonkman J, Castellarin SD, Bohlmann J, Page JE, Samuels AL (2020) Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. *Plant J* 101(1):37–56. <https://doi.org/10.1111/TPJ.14516>
- Luo X, Reiter MA, d’Espaux L, Wong J, Denby CM, Lechner A, Zhang Y, Grzybowski AT, Harth S, Lin W, Lee H, Yu C, Shin J, Deng K, Benites VT, Wang G, Baidoo EEK, Chen Y, Dev I, Petzold CJ, Keasling JD (2019) Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* 567(7746):123–126. <https://doi.org/10.1038/s41586-019-0978-9>
- Ma H, Wu Y, Lv R, Chi H, Zhao Y, Li Y, Liu H, Ma Y, Zhu L, Guo X, Kong J, Wu J, Xing C, Zhang X, Min L (2022) Cytochrome P450 mono-oxygenase CYP703A2 plays a central role in sporopollenin formation and ms5ms6 fertility in cotton. *J Integr Plant Biol* 64(10):2009–2025. <https://doi.org/10.1111/jipb.13340>
- MacWilliams J, Peirce E, Pitt WJ, Schreiner M, Matthews T, Yao L, Broeckling C, Nachappa P (2023) Assessing the adaptive role of cannabidiol (CBD) in *Cannabis sativa* defense against cannabis aphids. *Front Plant Sci* 14:12223894. <https://doi.org/10.3389/fpls.2023.12223894>
- Marks MD, Tian L, Wenger JP, Omburo SN, Soto-Fuentes W, He J, Gang DR, Weiblen GD, Dixon RA (2009) Identification of candidate genes affecting Δ^9 -tetrahydrocannabinol biosynthesis in *Cannabis sativa*. *J Exp Bot* 60(13):3715–3726. <https://doi.org/10.1093/jxb/erp210>
- McKernan KJ, Helbert Y, Kane LT, Ebling H, Zhang L, Liu B, Eaton Z, McLaughlin S, Kingan S, Baybayan P, Concepcion G, Jordan M, Riva A, Barbazuk W, Harkins T (2020) Sequence and annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes. *bioRxiv*. <https://doi.org/10.1101/2020.01.03.894428>
- McKernan K, Helbert Y, Kane LT, Ebling H, Zhang L, Liu B, Eaton Z, Sun L, Dimalanta ET, Kingan S, Baybayan P, Press M, Barbazuk W, Harkins T (2018) Cryptocurrencies and Zero Mode Wave guides: An unclouded path to a more contiguous *Cannabis sativa* L genome assembly. *OSF*. <https://doi.org/10.31219/osf.io/7d968>
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37(5):1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Nishikawa S, Zinkl GM, Swanson RJ, Maruyama D, Preuss D (2005) Callose (β -1,3 glucan) is essential for *Arabidopsis* pollen wall patterning, but not tube growth. *BMC Plant Biol* 5(1):22. <https://doi.org/10.1186/1471-2229-5-22>
- Pieracci Y, Ascrizzi R, Terreni V, Pistelli L, Flamini G, Bassolino L, Fulvio F, Montanari M, Paris R (2021) Essential oil of *Cannabis sativa* L: comparison of yield and chemical composition of 11 hemp genotypes. *Molecules* 26(13):4080. <https://doi.org/10.3390/molecules26134080>
- Rea KA, Casaretto JA, Al-Abdul-Wahid MS, Sukumaran A, Geddes-McAlister J, Rothstein SJ, Akhtar TA (2019) Biosynthesis of cannflavins A and B from *Cannabis sativa* L. *Phytochemistry* 164:162–171. <https://doi.org/10.1016/J.PHYTOCHEM.2019.05.009>
- Reinders A (2016) Fuel for the road – sugar transport and pollen tube growth. *J Exp Bot* 67(8):2121–2123. <https://doi.org/10.1093/jxb/erw113>
- Ren M, Tang Z, Wu X, Spengler R, Jiang H, Yang Y, Boivin N (2019) The origins of cannabis smoking: chemical residue evidence from the first millennium BCE in the Pamirs. *Sci Adv* 5(6):1391–1403. <https://doi.org/10.1126/sciadv.aaw1391>
- Rice S, Kozziel JA (2015) Characterizing the smell of marijuana by odor impact of volatile compounds: an application of simultaneous chemical and sensory analysis. *PLoS ONE* 10(12):e0144160. <https://doi.org/10.1371/journal.pone.0144160>
- Ristevski, N. (2023). *MADS Box Genes in Cannabis sativa: A Genomic and Phylogenetic Analysis*. La Trobe University.
- Ryu B-R, Gim G-J, Shin Y-R, Kang M-J, Kim M-J, Kwon T-H, Lim Y-S, Park S-H, Lim J-D (2024) Chromosome-level haploid assembly of *Cannabis sativa* L. cv. Pink Pepper. *Sci Data* 11(1):1442. <https://doi.org/10.1038/s41597-024-04288-8>
- Schilling S, Dowling CA, Shi J, Ryan L, Hunt DJL, O’Reilly E, Perry AS, Kinnane O, McCabe PF, Melzer R (2021) The cream of the crop: biology, breeding, and applications of *Cannabis sativa*. *Annual Plant Reviews online*. John Wiley & Sons, Ltd, pp 471–528. <https://doi.org/10.1002/9781119312994.apr0740>
- Shah L, Sohail A, Ahmad R, Cheng S, Cao L, Wu W (2022) The roles of MADS-Box genes from root growth to maturity in *Arabidopsis* and rice. *Agronomy* 12(3):Article 3. <https://doi.org/10.3390/agronomy12030582>
- Di Stilio VS, Kramer EM, Baum DA (2005) Floral MADS box genes and homeotic gender dimorphism in *Thalictrum dioicum* (*Ranunculaceae*) – a new model for the study of dioecy. *Plant J* 41(5):755–766. <https://doi.org/10.1111/j.1365-3113X.2005.02336.x>
- Thomas F, Schmidt C, Kayser O (2020) Bioengineering studies and pathway modeling of the heterologous biosynthesis of tetrahydrocannabinolic acid in yeast. *Appl Microbiol Biotechnol* 104(22):9551–9563. <https://doi.org/10.1007/s00253-020-10798-3>
- Tian F, Yang D-C, Meng Y-Q, Jin J, Gao G (2020) PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res* 48(D1):D1104–D1113. <https://doi.org/10.1093/nar/gkz1020>
- Usadel B, Schlüter U, Møhlhøj M, Gimpans M, Verma R, Kossmann J, Reiter W-D, Pauly M (2004) Identification and characterization of a UDP-d-glucuronate 4-epimerase in *Arabidopsis*. *FEBS Lett* 569(1):327–331. <https://doi.org/10.1016/j.febslet.2004.06.005>
- Welling M, Deseo M, O’Brien M, Clifton J, Bacic A, Doblin M (2023) Metabolomic analysis of methyl jasmonate treatment on phytocannabinoid production in *Cannabis sativa*. *Front Plant Sci* 14:1110144. <https://doi.org/10.3389/fpls.2023.1110144>
- Yadav S, Ghimire N, Lahutiya V (2023) A comprehensive review of the production technology of *Cannabis sativa* L. with its current legal status and botanical features. *Fundam Appl Agric* 8(1):458. <https://doi.org/10.5455/faa.144546>

- Yang X, Wu D, Shi J, He Y, Pinot F, Grausem B, Yin C, Zhu L, Chen M, Luo Z, Liang W, Zhang D (2014) Rice CYP703A3, a cytochrome P450 hydroxylase, is essential for development of anther cuticle and pollen exine. *J Integr Plant Biol* 56(10):979–994. <https://doi.org/10.1111/jipb.12212>
- Zhang C, Sultan SA, T R, Chen X (2021) Biotechnological applications of S-adenosyl-methionine-dependent methyltransferases for natural products biosynthesis and diversification. *Bioresour Bioprocess* 8(1):72. <https://doi.org/10.1186/s40643-021-00425-y>
- Zhang J, Yan J, Huang S, Pan G, Chang L, Li J, Zhang C, Tang H, Chen A, Peng D, Biswas A, Zhang C, Zhao L, Li D (2020) Genetic Diversity and Population Structure of Cannabis Based on the Genome-Wide Development of Simple Sequence Repeat Markers. *Front Genet* 11:958. <https://doi.org/10.3389/fgene.2020.00958>
- Rambaut, A. (2012). *FigTree* (Version 1.4.4) [Computer software]. <http://tree.bio.ed.ac.uk/software/figtree/>
- Releaf. (2023, July 6). *Exploring the anatomy of the cannabis flower*. <https://releaf.co.uk/education/cannabis-101/medical-cannabis/exploring-the-anatomy-of-the-cannabis-flower>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Kevelin Barbosa-Xavier¹  · Thiago M. Venancio¹ 

✉ Thiago M. Venancio
tmvenancio@uenf.br

Lamego, 2000, Parque Califórnia, Campos dos Goytacazes,
RJ CEP 28013-602, Brazil

¹ Laboratório de Química e Função de Proteínas e Peptídeos,
Centro de Biociências e Biotecnologia, Universidade
Estadual do Norte Fluminense Darcy Ribeiro, Av. Alberto