



Genome-Wide Characterization of the *MLO* Gene Family in *Cannabis sativa* Reveals Two Genes as Strong Candidates for Powdery Mildew Susceptibility

Noémi Pépin^{1†}, Francois Olivier Hebert^{1,2†} and David L. Joly^{1*}

¹ Centre d'Innovation et de Recherche sur le Cannabis, Université de Moncton, Département de biologie, Moncton, NB, Canada, ² Institut National des Cannabinoïdes, Montréal, QC, Canada

OPEN ACCESS

Edited by:

Donald Lawrence Smith,
McGill University, Canada

Reviewed by:

Michele Wiseman,
Oregon State University,
United States
Kevin McKernan,
Medicinal Genomics Corporation,
United States

*Correspondence:

David L. Joly
david.joly@umoncton.ca

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Symbiotic Interactions,
a section of the journal
Frontiers in Plant Science

Received: 22 June 2021

Accepted: 19 August 2021

Published: 13 September 2021

Citation:

Pépin N, Hebert FO and Joly DL
(2021) Genome-Wide
Characterization of the *MLO* Gene
Family in *Cannabis sativa* Reveals
Two Genes as Strong Candidates
for Powdery Mildew Susceptibility.
Front. Plant Sci. 12:729261.
doi: 10.3389/fpls.2021.729261

Cannabis sativa is increasingly being grown around the world for medicinal, industrial, and recreational purposes. As in all cultivated plants, cannabis is exposed to a wide range of pathogens, including powdery mildew (PM). This fungal disease stresses cannabis plants and reduces flower bud quality, resulting in significant economic losses for licensed producers. The *Mildew Locus O (MLO)* gene family encodes plant-specific proteins distributed among conserved clades, of which clades IV and V are known to be involved in susceptibility to PM in monocots and dicots, respectively. In several studies, the inactivation of those genes resulted in durable resistance to the disease. In this study, we identified and characterized the *MLO* gene family members in five different cannabis genomes. Fifteen *Cannabis sativa MLO (CsMLO)* genes were manually curated in cannabis, with numbers varying between 14, 17, 19, 18, and 18 for CBDRx, Jamaican Lion female, Jamaican Lion male, Purple Kush, and Finola, respectively (when considering paralogs and incomplete genes). Further analysis of the *CsMLO* genes and their deduced protein sequences revealed that many characteristics of the gene family, such as the presence of seven transmembrane domains, the *MLO* functional domain, and particular amino acid positions, were present and well conserved. Phylogenetic analysis of the *MLO* protein sequences from all five cannabis genomes and other plant species indicated seven distinct clades (I through VII), as reported in other crops. Expression analysis revealed that the *CsMLOs* from clade V, *CsMLO1* and *CsMLO4*, were significantly upregulated following *Golovinomyces ambrosiae* infection, providing preliminary evidence that they could be involved in PM susceptibility. Finally, the examination of variation within *CsMLO1* and *CsMLO4* in 32 cannabis cultivars revealed several amino acid changes, which could affect their function. Altogether, cannabis *MLO* genes were identified and characterized, among which candidates potentially involved in PM susceptibility were noted. The results of this study will lay the foundation for further investigations, such as the functional characterization of clade V *MLOs* as well as the potential impact of the amino acid changes reported. Those will be useful for breeding purposes in order to develop resistant cultivars.

Keywords: *Cannabis*, powdery mildew, fungal disease, susceptibility genes, plant–pathogen interactions, *MLO*

INTRODUCTION

Cannabis sativa is a dicotyledonous plant belonging to the Cannabaceae family, and it is considered a socially and economically important crop as it is increasingly being grown and cultivated around the world. In Canada alone, the sales from cannabis stores in 2020 reached over 2.6 billion dollars (Statistics Canada, 2021), and the number of licensed cultivators, processors, and sellers quadrupled from 2018 to 2020 (Health Canada, 2020). It is used as a source of industrial fiber, seed oil, food, as well as for medicinal, spiritual, and recreational purposes (Small, 2015). As in all cultivated plants, cannabis is exposed to numerous pathogens, and the resulting diseases play a limiting role in its production.

Powdery mildew (PM) is a widespread plant disease caused by ascomycete fungi of the order Erysiphales, for which more than 800 species have been described (Braun and Cook, 2012). They are obligate biotrophs that form invasive structures in epidermal cells for nutrient uptake, called haustoria (Glawe, 2008). These pathogens can infect nearly 10,000 monocotyledonous and dicotyledonous plant species and cause significant damage to crops and ornamental plants (Braun et al., 2002). The PM disease in cannabis, caused by *Golovinomyces ambrosiae* emend. (including *Golovinomyces spadiceus*), has been reported on indoor- and greenhouse-grown plants in Canada and in the United States, where the enclosed conditions provide an ideal environment for the germination and propagation of the fungal spores (Pépin et al., 2018; Szarka et al., 2019; Farinas and Peduto Hand, 2020; Weldon et al., 2020; Wiseman et al., 2021). An analysis of cannabis buds revealed *Golovinomyces* sp. in 79% of tested samples (Thompson et al., 2017), highlighting its ubiquity among licensed producers. The symptoms initially appear as white patches on leaves, and eventually, the mycelia progress to cover the entire leaf surface, the flower bracts, and buds, resulting in stressed and weakened plants, reduced yield, and reduced flower buds quality. Fungicides are widely used to prevent and control this disease in agricultural settings. However, a scarce amount of such products are currently approved by Health Canada as the presence of fungicide residues in the inflorescences raises concerns (Punja, 2021). Besides, they are costly, and fungicide resistance in PM has been observed and documented in other plant species in recent years (Vielba-Fernández et al., 2020). Alternative approaches to managing this disease have been described, such as the use of biological control (e.g., *Bacillus subtilis* strain QST 713), reduced risk products (e.g., potassium bicarbonate, knotweed extract), and physical methods (e.g., de-leafing, irradiation) (Punja, 2021). Nonetheless, some of these methods increase production costs, are labor-intensive, and necessitate further research. Therefore, identifying sources of genetic resistance to PM in cannabis and ultimately breeding or developing resistant cultivars offer the most effective and sustainable approach to controlling PM.

A common strategy used in resistance breeding relies on the exploitation of resistance genes in plants, which encodes for cytoplasmic receptors such as nucleotide-binding leucine-rich repeat proteins or surface receptors such as receptor-like kinases and receptor-like proteins. These immune receptors can

detect specific proteins or molecules produced by the pathogen and subsequently induce plant defense responses (Dangl and Jones, 2001). While useful, most resistance genes confer race-specific resistance and are therefore frequently overcome by the emergence of a pathogen's new virulent race within a few years. An alternative approach in resistance breeding is to exploit susceptibility genes (*S-genes*) in plants. *S-genes* are defined as genes that facilitate infection and support compatibility for a pathogen (van Schie and Takken, 2014). The alteration of such genes can limit the pathogen's ability to infect the plant and therefore provide a durable type of resistance (van Schie and Takken, 2014). Such PM resistance was initially observed in an X-ray irradiated barley (*Hordeum vulgare*) population in the 1940s (Freisleben and Lein, 1942). It was discovered later that the immunity was attributable to a mutated *S-gene* named *Mildew Locus O (MLO)*, which was recessively inherited. Complete resistance to all known isolates of PM, caused by *Blumeria graminis* f. sp. *hordei* was conferred in barley by these loss-of-function mutations when present in the homozygous state. This type of resistance has been durable under field conditions and has been used for over 40 years in barley breeding programs, without any break in the resistance (Jørgensen, 1992; Büschges et al., 1997; Lyngkjær and Carver, 2000; Piffanelli et al., 2002).

Since discovering *MLO* genes in barley, many *MLO* homologs have been identified in several plant species, especially in monocots and eudicots, as the PM disease affects angiosperms solely. For instance, *MLO* genes were identified in Rosaceae [roses (Kaufmann et al., 2012), apple, peach, strawberry and apricot (Pessina et al., 2014)], Cucurbitaceae [cucumber (Zhou et al., 2013), melon, watermelon, zucchini (Iovieno et al., 2015) and pumpkin (Win et al., 2018)], Solanaceae [tomato (Bai et al., 2008), pepper (Zheng et al., 2013), tobacco, potato and eggplant (Appiano et al., 2015)], Fabaceae [pea (Humphry et al., 2011; Pavan et al., 2011), soybean (Shen et al., 2012; Deshmukh et al., 2014), barrel medic, chickpea, narrow-leaf lupin, peanut, pigeon pea, common bean, mungbean (Rispaill and Rubiales, 2016) and lentil (Polanco et al., 2018)], Brassicaceae [thale cress (Devoto et al., 1999, 2003)], Vitaceae [grapevine (Feechan et al., 2008)], and Poaceae [rice (Liu and Zhu, 2008), wheat (Konishi et al., 2010), sorghum (Singh et al., 2012), maize (Devoto et al., 2003; Kusch et al., 2016), foxtail millet (Kusch et al., 2016) and stiff brome (Ablazov and Tombuloglu, 2016)]. Furthermore, thorough phylogenetic analyses of land plants revealed that *MLO* genes were not only present in monocots and eudicots but also in basal angiosperms, gymnosperms, lycophytes, and bryophytes (Jiwan et al., 2013; Kusch et al., 2016; Shi et al., 2020). Many *MLO*-like proteins were also identified in algae and other unicellular eukaryotes, suggesting that *MLO* is an ancient eukaryotic protein (Jiwan et al., 2013; Kusch et al., 2016; Shi et al., 2020).

The *MLO* gene family is described as a medium-sized plant-specific gene family, with a varying number of members between 7 in wheat to 39 in soybean, depending on the species (Acevedo-García et al., 2014). The resulting *MLO* proteins are characterized by the presence of seven transmembrane domains integral to the plasma membrane with an extracellular N-terminus and an intracellular C-terminus (Devoto et al., 1999). They are also characterized by the presence of a calmodulin-binding domain

in the C-terminal region that is likely implicated in sensing calcium influx and mediating various signaling cascades (Kim et al., 2002a,b). MLO protein sequences identified across land plants also possess several highly conserved amino acids, some of which have been deemed essential for the structure, functionality, and stability of the protein (Devoto et al., 2003; Elliott et al., 2005; Reinstädler et al., 2010; Kusch et al., 2016). Mutations in these residues could affect the accumulation, maturation, and function of the protein and are therefore attractive targets for breeding programs (Elliott et al., 2005; Reinstädler et al., 2010).

Throughout land plant evolution, the MLO protein family diversified into subfamilies, or clades, which have been demonstrated in several phylogenetic analyses. MLO proteins are usually grouped into seven defined clades (I to VII), among which clades IV and V appear to host MLO proteins associated with PM susceptibility in monocots and dicots, respectively (Acevedo-Garcia et al., 2014; Kusch et al., 2016). It has been documented in many species, such as barley, tomato, and apple, that MLO genes from these two clades (IV and V in monocots and dicots, respectively) are up-regulated upon PM infection (Piffanelli et al., 2002; Zheng et al., 2013; Pessina et al., 2014). It has also been demonstrated that the overexpression of these genes results in enhanced susceptibility to the pathogen (Zheng et al., 2013). Furthermore, the inactivation of these genes in many species by gene silencing, genome editing, or TILLING has resulted in increased or complete resistance to PM (Wang et al., 2014; Acevedo-Garcia et al., 2017; Nekrasov et al., 2017; Ingvarsdson et al., 2019; Wan et al., 2020). Besides the implication of clade V and IV MLOs in PM susceptibility, recent studies have suggested that MLO genes from other clades are implicated in various physiological and developmental processes. For example, it was demonstrated that in *Arabidopsis thaliana*, *AtMLO4* and *AtMLO11* from clade I are involved in root thigmomorphogenesis (Chen et al., 2009; Bidzinski et al., 2014), while *AtMLO7* from clade III is involved in pollen tube reception by the embryo sac (Kessler et al., 2010). In rice (*Oryza sativa*), *OsMLO12* from clade III mediates pollen hydration (Yi et al., 2014). Interestingly, barley *HvMLO1* has been shown to differentially regulate the establishment of mutualistic interactions with the endophyte *Serendipita indica* and the arbuscular mycorrhizal fungus *Funneliformis mosseae* (Hilbert et al., 2020). Indeed, another study clearly showed barley *HvMLO1*, wheat *TaMLO1*, and barrel medic *MtMLO8* from clade IV to be involved in the establishment of symbiotic relationships with beneficial mycorrhizal fungi (Jacott et al., 2020). These findings suggest that PMs might have appropriated and exploited these genes as an entryway to successful pathogenic colonization (Jacott et al., 2020). However, in pea, no evidence was found for the implication of *PsMLO1*, a clade V gene, in the establishment of relationships with mycorrhizal and rhizobial symbionts (Humphry et al., 2011).

Although MLO genes have been studied in many monocot and dicot species, they have only been preliminarily studied in cannabis (McKernan et al., 2020). The growing interest in cannabis research has led to the publication of several genomes in recent years, thus providing an opportunity to conduct a comprehensive analysis of the MLO gene family in cannabis.

In this study, we manually curated and characterized the members of the MLO gene family in cannabis from five different available genomes: Purple Kush and Finola (Lavery et al., 2019), CBDRx (Grassa et al., 2021), and Jamaican Lion (female, McKernan et al., 2018; male, McKernan et al., 2020). Through phylogenetic analysis, we identified candidate MLO genes likely to be involved in PM susceptibility in cannabis, observed their subcellular localization by confocal microscopy, and monitored their expression profile in cannabis leaves during infection. We also searched for potential naturally occurring resistant mutants by investigating amino acid changes in 32 cultivars. A better understanding of cannabis MLOs offers enormous opportunities to breed PM-resistant cultivars and develop new control methods, thereby increasing productivity and yield.

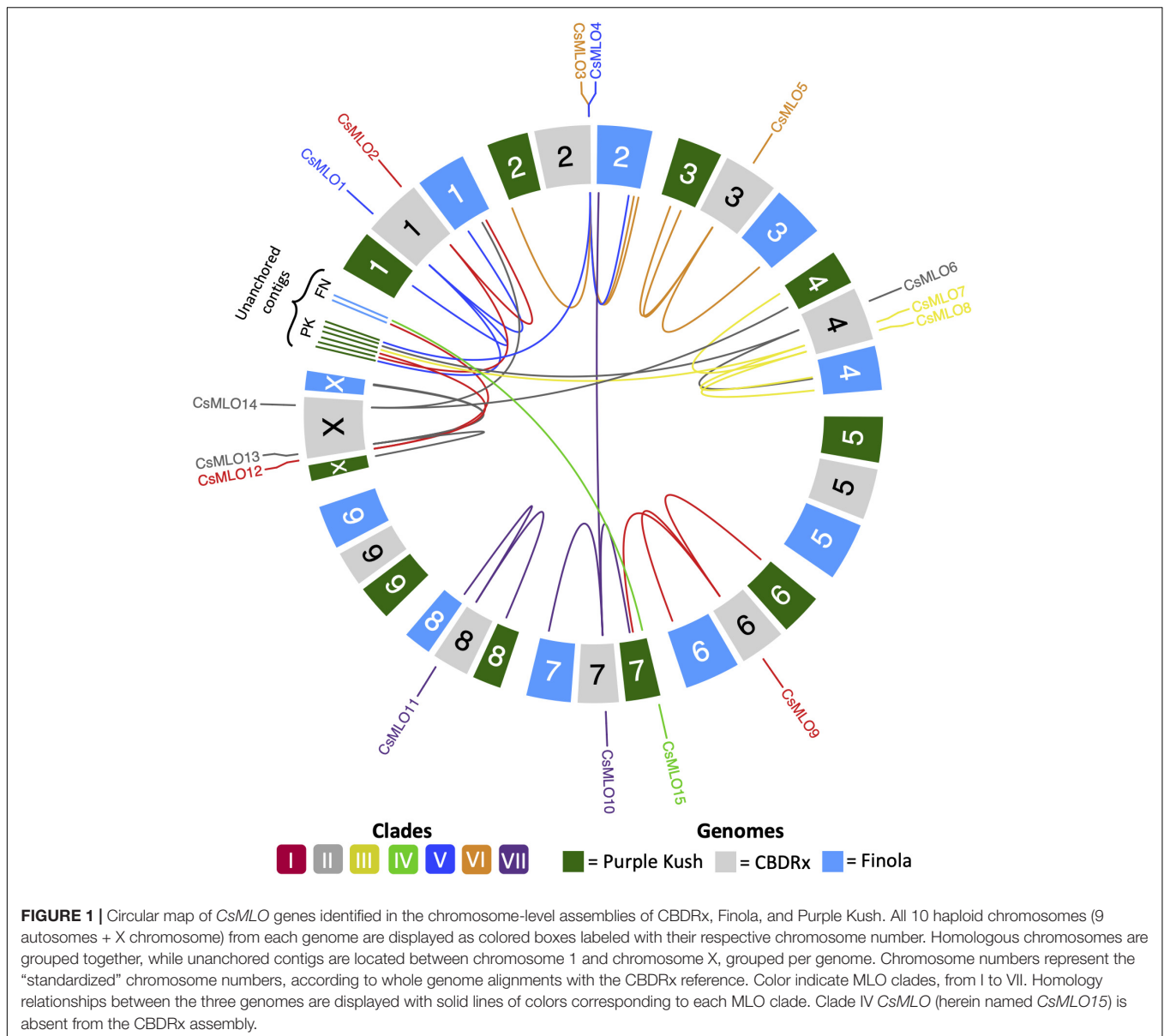
MATERIALS AND METHODS

In silico Identification and Manual Curation of the Cannabis MLO Genes

Cannabis MLO genes in CBDRx were initially identified (also named 'cs10,' NCBI accession GCA_900626175.2) using TBLASTN from the BLAST+ suite (Camacho et al., 2009) with *Arabidopsis thaliana* amino acid sequences as queries (*AtMLO1-AtMLO15*, NCBI accession numbers: NP_192169.1, NP_172598.1, NP_566879.1, NP_563882.1, NP_180923.1, NP_176350.1, NP_179335.3, NP_565416.1, NP_174980.3, NP_201398.1, NP_200187.1, NP_565902.1, NP_567697.1, NP_564257.1, NP_181939.1). In parallel, all official gene models were extracted from the NCBI CBDRx annotation report (NCBI Annotation Release 100¹) with an InterPro (IPR004326) and/or Pfam (PF03094) identification number related to the MLO gene family (Blum et al., 2021; Mistry et al., 2021). These two sequence datasets were merged together, and multiple sequence alignments were performed using MUSCLE v.3.8.31 (Edgar, 2004) with the genomic and the amino acid versions of each MLO gene model. Only unique sequences were kept, and all MLO gene models that were retained but incomplete were further manually curated. The full genomic sequences were aligned and manually compared with their corresponding full-length mRNA transcripts using BLASTN from the BLAST+ suite (Camacho et al., 2009). Each gene was characterized based on total length, chromosome localization, strand, START and STOP positions, as well as number and size of exons and introns (Figure 1 and Table 1). The resulting streamlined and manually curated MLO gene models for CBDRx were considered the definitive reference set for this genome.

The reference set of CBDRx MLO genes were used as queries to search for the presence of homologs in four other cannabis genomes (Purple Kush – GCA_000230575.5, Finola – GCA_003417725.2, Jamaican Lion female – GCA_012923435.1, Jamaican Lion male – GCA_013030025.1), using TBLASTN. Manual curation of each set of MLO genes was performed in each of these genomes, using the same approach described previously for CBDRx. In the process, several frameshifts (mostly small

¹https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Cannabis_sativa/100/



insertions and deletions) were noted and manually corrected in the coding sequence of multiple *MLOs* in Finola (13 genes) and Purple Kush (7 genes). None of the homologs for these genes showed any frameshift in any other genome. All the frameshifts identified in Finola and Purple Kush were thus examined by comparing their respective coding sequence with available transcriptomic data from the CanSat3 assembly project², using BLASTN. Based on evidence from mRNA sequences, all these frameshifts were manually corrected. All of the manually curated *MLO* genes for all five cannabis genomes were ultimately considered as our reference and final set of *Cannabis sativa* *MLO* (*CsMLO*) genes (Table 1, Supplementary Tables 1–5, and Supplementary Files 1–3). The structure of our final *CsMLO* gene models was compared to their respective genome

annotation available on NCBI (CBDRx, Jamaican Lion male and female) (Supplementary Tables 1–3).

Chromosomal localization of each manually curated *CsMLO* gene in the chromosome-level genome assemblies, i.e., CBDRx, Finola, and Purple Kush, was displayed on a chromosomal map (Figure 2) with the R package Circlize (Gu et al., 2014). As chromosome information is not available for Jamaican Lion female and Jamaican Lion male, those were not considered for this analysis. Chromosome numbers in Finola and Purple Kush were standardized to match the official chromosome numbers in the CBDRx reference genome, using whole genome alignments on the D-Genies platform³. This analysis served as a preliminary assessment of synteny for *CsMLO* genes.

²<http://genome.ccbcr.utoronto.ca/downloads.html>

³<http://dgenies.toulouse.inra.fr/>

TABLE 1 | Members of the *CsMLO* gene family as predicted and manually curated in the *CBDRx* genome.

Gene	Clade	Chr.	Strand	Start position	End position	Genomic length (bp)	Exons	Protein length (aa)	TMs ^b	Subcellular localization ^c	Conserved aa/30 ^d	Conserved aa/58 ^e
<i>CsMLO1-CBDRx</i>	V	1	+	21,631,499	21,636,477	4,979	15	555	7	Cell membrane	28	57
<i>CsMLO2-CBDRx</i>	I	1	-	84,421,095	84,415,528	5,568	15	551	7	Cell membrane	30	58
<i>CsMLO3-CBDRx</i>	VI	2	-	92,590,377	92,584,053	6,325	15	547	7	Cell membrane	25	52
<i>CsMLO4-CBDRx</i>	V	2	+	92,612,089	92,622,483	10,395	15	631	7	Cell membrane	30	56
<i>CsMLO5-CBDRx</i>	VI	3	-	42,556,462	42,507,454	49,009	15	515	7	Cell membrane	30	58
<i>CsMLO6-CBDRx</i>	II	4	-	29,809,460	29,799,903	9,558	14	520	7	Cell membrane	30	57
<i>CsMLO7-CBDRx</i>	III	4	-	65,440,972	65,435,678	5,295	15	549	7	Cell membrane	29	57
<i>CsMLO8-CBDRx</i>	III	4	+	79,070,668	79,074,427	3,760	15	542	7	Cell membrane	30	57
<i>CsMLO9-CBDRx</i>	I	6	+	49,150,225	49,156,766	6,542	15	568	7	Cell membrane	30	57
<i>CsMLO10-CBDRx</i>	VII	7	-	23,870,119	23,830,282	39,838	14	535	7	Cell membrane	30	57
<i>CsMLO11-CBDRx</i>	VII	8	-	51,420,983	51,416,370	4,614	14	583	7	Cell membrane	30	58
<i>CsMLO12-CBDRx</i>	I	X	+	443,359	447,357	3,999	14	558	7	Cell membrane	29	55
<i>CsMLO13-CBDRx</i>	II	X	+	10,840,654	10,844,220	3,567	14	508	7	Cell membrane	30	58
<i>CsMLO14-CBDRx</i>	II	X	+	89,459,459	89,463,372	3,914	12	516	7	Cell membrane	30	56
<i>CsMLO15-CBDRx</i> ^a	IV	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

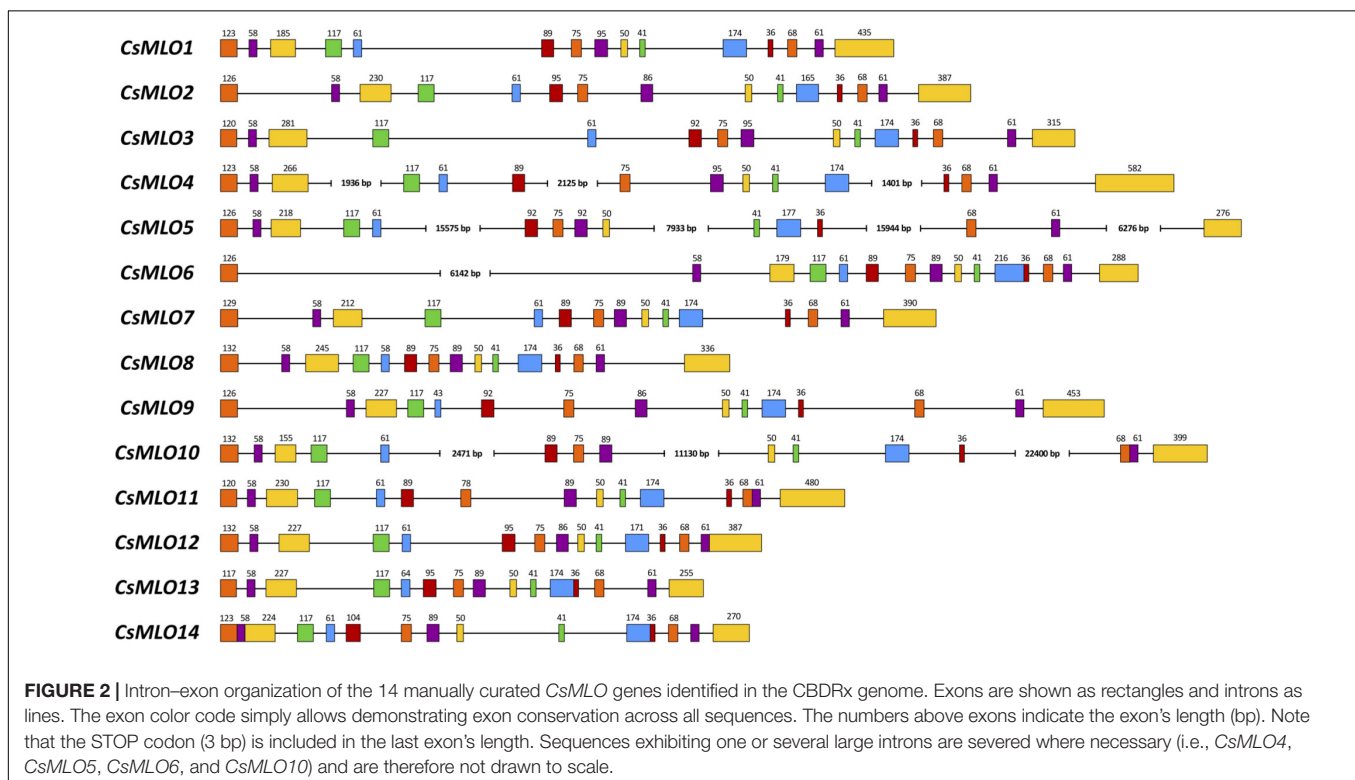
^a*CsMLO15-CBDRx* is not predicted in the *CBDRx* genome, as compared to the other four genomes analyzed.

^bNumber of transmembrane domains (TMs) in the predicted protein, as determined by the CCTOP online prediction server (Dobson et al., 2015).

^cSubcellular localization as predicted by three online prediction servers: Plant-mPLOC (Chou and Shen, 2010), YLoc-HighRes (Briesemeister et al., 2010), and DeepLoc 1.0 (Almagro Armenteros et al., 2017).

^dNumber of conserved amino acids out of the 30 identified in Elliott et al. (2005).

^eNumber of conserved amino acids out of the 58 identified in Kusch et al. (2016).



Gene and Protein Characterization

The protein sequences from the five genomes were analyzed through several online prediction servers in order to identify functional domains (InterProScan, Jones et al., 2014), transmembrane domains (CCTOP, Dobson et al., 2015),

subcellular localizations (Plant-mPLOC, Chou and Shen, 2010; DeepLoc 1.0, Almagro Armenteros et al., 2017; YLoc-HighRes, Briesemeister et al., 2010), signal peptide (SignalP 5.0, Almagro Armenteros et al., 2019a), calmodulin-binding domains (CaMELS, Abbasi et al., 2017) as well as mitochondrial,

chloroplast, and thylakoid luminal transit peptide (TargetP 2.0, Almagro Armenteros et al., 2019b). Conserved amino acids described in Elliott et al., 2005 (30 invariant amino acids) and Kusch et al., 2016 (58 highly conserved amino acids) were screened in all *CsMLO* sequences using our final protein alignment (**Supplementary File 3**). Our manually curated *MLO* gene models in the five cannabis genomes were also screened for conserved *cis*-acting elements in the promoter regions. A homemade Python script (v.3.7.3, Van Rossum and Drake, 2009) was used to extract a 2 kb upstream region of each *CsMLO* gene, and these promoter regions were used as search queries in the plantCARE database (Lescot et al., 2002).

Phylogenetic Analysis

Amino acid sequences for all manually curated *CsMLOs* from CBDRx, Jamaican Lion female and male, Purple Kush and Finola were aligned together with *MLO* sequences previously identified in *Arabidopsis thaliana* (AtMLOs, as indicated in Devoto et al., 2003), *Vitis vinifera* (grapevine: VvMLOs, as indicated in Feechan et al., 2008), *Prunus persica* (peach: PpMLOs, as indicated in Pessina et al., 2014), *Hordeum vulgare* (barley: HvMLOs, as indicated in Kusch et al., 2016) and *Zea mays* (maize: ZmMLOs, as indicated in Kusch et al., 2016). *Chlamydomonas reinhardtii* MLO (XP_001689918) was used as an outgroup. Alignment of protein sequences was performed using MAFFT v7.407_1 (Kato and Standley, 2013) with default parameters within NGPhylogeny.fr (Lemoine et al., 2019) and used to construct phylogenetic trees. A first tree was constructed using PhyML+SMS v1.8.1_1 (Lefort et al., 2017) with default parameters within NGPhylogeny.fr (**Figure 3**), and a second tree was constructed using MrBayes v3.2.6_1 (Huelsenbeck and Ronquist, 2001) with default parameters within NGPhylogeny.fr (**Supplementary Figure 1**). PhyML+SMS implements SMS (Smart Model Selection) that uses a heuristic approach for model selection. The trees were interpreted and visualized using iTOL (Letunic and Bork, 2019). All *MLO* proteins identified were classified into clades based on previous phylogenetic analyses (Kusch et al., 2016).

Transcriptional Activity of *CsMLOs* in Response to Powdery Mildew Infection Sampling and RNA Sequencing

An RNA-seq time series analysis of the infection of cannabis by PM was performed to characterize the transcriptional response of *CsMLO* genes. The experiment was conducted in a controlled environment at Organigram Inc., a Health Canada approved licensed producer (Moncton, New-Brunswick, Canada). Cannabis fan leaves from 4-week-old vegetative plants ('Pineapple Express,' drug-type I) were manually inoculated with *G. ambrosiae* emend. (including *G. spadicus*) spores. Heavily infected leaves, loaded with fungal spores, were scraped against the surface of the leaves from the 4-week-old plants to induce infection. Leaf samples (punch holes) were taken at five time points during the infection ($n = 3$ per time point): day zero (T0), 6 h post-inoculation (6H), 1 day post-inoculation (1D), 3 days post-inoculation (3D) and 8 days post-inoculation (8D). RNA

samples were extracted from leaf tissues using the RNeasy® Plant Mini Kit (QIAGEN, Hilden, Germany) following the standard manufacturer's protocol. Each mRNA extraction was treated with QIAGEN's RNase-Free DNase Set (QIAGEN, Hilden, Germany), involving a first round of the TURBO DNA-free™ DNA Removal Kit through the extraction protocol and then two rounds of the DNA-free™ Kit (Life Technologies, Carlsbad, CA, United States). Fifteen individual RNA-seq libraries were generated for the five time points sampled ($n = 3$ per time point). cDNA libraries were sequenced on a total of six lanes, using the Illumina HiSeq v4 technology (PE 125 bp) at the Centre d'expertise et de services Génome Québec (Montreal, QC, Canada). In total, ~200 Gb of raw sequencing data were generated, which represents 1.728 billion of 2×125 bp paired-end sequences distributed across all 15 libraries (BioProject accession: PRJNA738505, SRA accessions: SRR14839036-50).

Short-Read Alignment on the Reference Genome and Differential Gene Expression

Raw sequencing reads were cleaned, trimmed, and aligned on the same reference genome as the one initially used to find our final *MLO* gene models (CBDRx) to estimate changes in transcript-specific levels of expression over the course of the infection. Specifically, mild trimming thresholds were applied to clean and trim all raw reads, using Trimmomatic v.0.34 (Bolger et al., 2014) with the following parameters: ILLUMINACLIP:\$VECTORS:2:30:10, SLIDINGWINDOW:20:2, LEADING:2, TRAILING:2, MINLEN:60. The cleaned reads from our 15 individual libraries were aligned on the CBDRx reference genome using STAR v.2.7.6 (Dobin et al., 2013) in genome mode with default parameters and the official NCBI Cannabis annotation release 100 associated with the genome. Genome-wide raw read counts were obtained for each library using htseq-count v.0.11.1 (Anders et al., 2015) with the 'intersection-non-empty' mode.

Downstream analyses of differential gene expression patterns were conducted using the R packages 'limma-voom' (Law et al., 2014) and edgeR (Robinson et al., 2010) in RStudio v.1.3.1073 (RStudio Team, 2020). Reference sequences with insufficient sequencing depth were filtered out by keeping only the ones with more than five Counts Per Million (CPM) in at least three samples. This mild CPM threshold allowed the filtering of very low coverage genes without losing too much information in the dataset. This filtered dataset was normalized using the Trimmed Mean of M-values (TMM) method implemented in edgeR. A transformation of the data was then performed using the 'limma-voom' function, which estimates the mean-variance relationship for each transcript, allowing for better and more robust comparisons of gene expression patterns across RNA-seq libraries (Law et al., 2014). Each transcript was finally fitted to an independent linear model with \log_2 (CPM) values as the response variable and the time point (0, 6 h, 24 h, 3 days, and 8 days post-infection) as the explanatory variable. All linear models were treated with limma's empirical Bayes analysis pipeline (Law et al., 2014). Differentially expressed genes were chosen based on a False Discovery Rate (FDR, Benjamini-Hochberg procedure) < 0.05 . Genomic regions corresponding to the curated CBDRx *MLO*

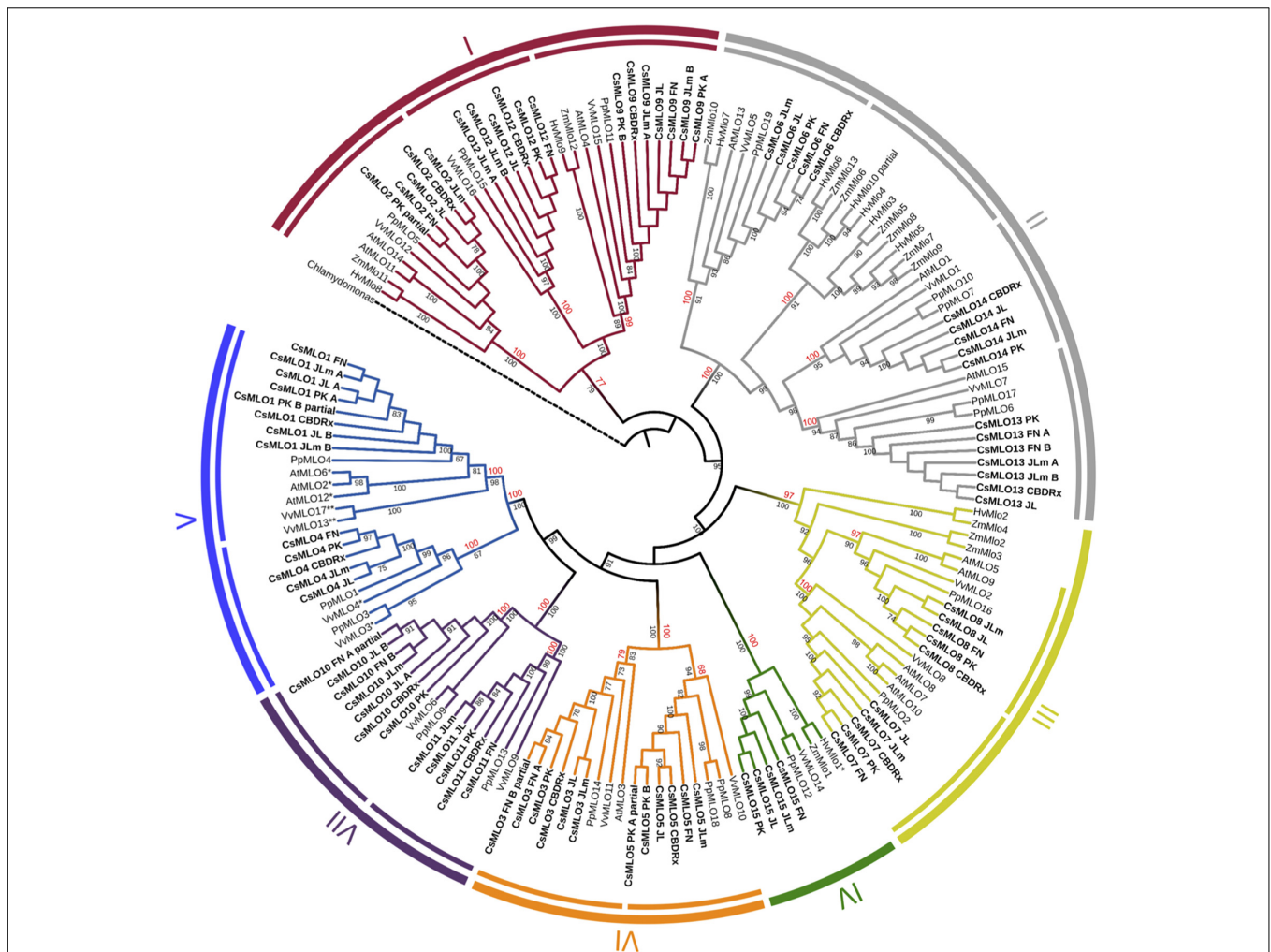


FIGURE 3 | Phylogenetic relationships of CsMLOs based on maximum likelihood analysis. Phylogenetic tree of manually curated CsMLO proteins (bold) with MLO proteins from selected species (*Arabidopsis thaliana*, *Prunus persica*, *Vitis vinifera*, *Hordeum vulgare*, and *Zea mays*). *Chlamydomonas reinhardtii* was used as an outgroup. Phylogenetic relationships were estimated using the maximum likelihood method implemented in PhyML + SMS with 1,000 bootstrap independent replicates. The seven defined clades were indicated, as well as potential subclades identified in this study (inner circles). Number at a node indicates the percentage of bootstrap when higher than 65% (black), or the posterior probabilities of major clades and subclades, according to a Bayesian phylogenetic inference performed on the same alignment (red) (Supplementary Figure 1). MLOs with one asterisk (*) have been experimentally demonstrated to be required for PM susceptibility (Büschges et al., 1997; Feechan et al., 2008; Wan et al., 2020), while MLOs with two asterisks (**) have been identified as main probable candidates for PM susceptibility (Pessina et al., 2016).

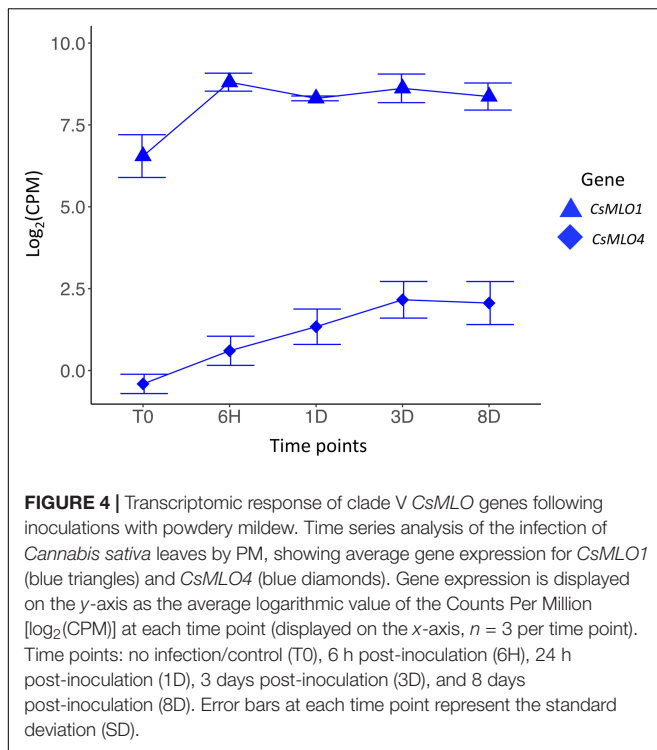
gene models were extracted from the output of edgeR/limma-voom for each comparison made (i.e., each time point compared to T0) and looked for significant gene expression differences among *MLO* genes ($FDR < 0.05$). These expression differences in *MLO*s over the course of the infection were visualized in RStudio v.1.3.1073 (RStudio Team, 2020) on a scatter plot using CPM values (Figure 4 and Supplementary Figure 2).

Cloning of Clade V CsMLOs for Transient Expression in *Nicotiana benthamiana* and Confocal Microscopy

The two selected *MLO* gene sequences (*CsMLO1* and *CsMLO4*) were synthesized commercially into the Gateway-compatible

vector pDONRTM/Zeo (Invitrogen, Carlsbad, CA, United States) (Hartley et al., 2000). The entry vectors were inserted into *Escherichia coli* OneShot[®] TOP10 cells (Invitrogen, Carlsbad, CA, United States) by chemical transformation according to the manufacturer's instructions. Positive colonies were selected, and plasmid DNA was extracted with the EZ-10 Spin Column Plasmid DNA Miniprep Kit (Bio Basic Inc., Markham, ON, Canada). The extracted entry vectors were confirmed by PCR with the primers M13-F (5'-GTAAAACGACGGCCAGT-3') and M13-R (5'-CAGGAAACAGCTATGAC-3') as well as M13-F and MLO1_1-R (5'-ATGTGCCATTATAAATCCATGCCT-3', this study).

The Gateway-compatible destination vector chosen was pB7FWG2.0, which is under the regulation of the 35S Cauliflower



Mosaic Virus (CaMV) promoter and harbors the plant selectable marker gene *bar* (bialaphos acetyltransferase), which confer resistance against glufosinate ammonium. It also possesses a streptomycin and spectinomycin resistance gene for plasmid selection, and an EGFP-fusion in C-terminal, for visualization by confocal microscopy (Karimi et al., 2002). According to the manufacturer's instructions, the destination vector was inserted into *Escherichia coli* One Shot[®] ccdB Survival[™] cells (Life Technologies) by chemical transformation. Positive colonies were selected, and plasmid DNA was extracted with the EZ-10 Spin Column Plasmid DNA Minipreps Kit (Bio Basic Inc., Markham, ON, Canada). The extracted destination vectors were confirmed by PCR with the primers T-35S-F (5'-AGGGTTTCTATATGCTCAACACATG-3', Debode et al., 2013) and EGFP-C (5'-CATGGTCCTGCTGGAGTTCGTG-3').

The two synthesized MLO gene sequences were then inserted into the vector pB7FWG2.0 through an LR clonase reaction following the manufacturer's instructions (Invitrogen, Carlsbad, CA, United States). Plasmids were then transferred to *E. coli* OneShot[®] TOP10 cells (Invitrogen, Carlsbad, CA, United States) by chemical transformation according to the manufacturer's instructions. Positive colonies were selected, and plasmid DNA was extracted with the EZ-10 Spin Column Plasmid DNA Minipreps Kit (Bio Basic Inc., Markham, ON, Canada). The extracted expression vectors were amplified by PCR with the primers 35S Promoter (5'-CTATCCTTCGCAAGACCCCTTC-3') and MLO1_1-R as well as MLO2_3-F (5'-TCTTTTCAGAATGCATTTCAACTTGC-3', this study) and EGFP-N (5'-CGTCGCCGTCCAGCTCGACCAG-3'). Sequencing of the PCR products above was performed to confirm

the junction between the plasmid and the inserted gene and the junction between the inserted gene and the GFP, respectively.

The recombinant vectors were then transferred to *Agrobacterium tumefaciens* ElectroMAX[™] LBA4404 cells by electroporation (Invitrogen, Carlsbad, CA, United States). The expression vector in *A. tumefaciens* was confirmed by PCR using the primers 35S Promoter and MLO1_1R and MLO2_3-F + EGFP-N. Cultures of transformed *A. tumefaciens* (*CsMLO1* and *CsMLO4*) were incubated with agitation at 28°C in LB broth containing spectinomycin (100 mg/mL) for 24 h. The cultures were then centrifuged at $5,000 \times g$ for 5 min, the supernatant was discarded, and the pellet was resuspended in MgCl_2 (10 mM). The cells were brought to an OD_{600} of 0.5 and incubated at room temperature for 2 h with acetosyringone (200 μM). *Nicotiana benthamiana* plants, about 2 weeks old, were watered a few hours before infiltration, and the bacterial suspensions were administered using sterile 1 mL syringes (without needles) on the abaxial surface of the leaves. The plants were then returned to growth chambers, and the observation of epidermal cells was performed 3 days after infiltration using a confocal laser scanning microscope. Leaves were observed under a Leica TCS SP8 confocal laser scanning microscopy (Leica Microsystems). Images were observed through an HC PL APO CS2 40X/1.40 oil immersion objective at excitation/emission wavelengths of 488/503–521 nm.

***In silico* Screening of Clade V *CsMLO* Sequence Variants in 32 Cannabis Cultivars**

Thirty-one distinct drug-type *Cannabis sativa* cultivars from Organigram Inc. (Moncton, NB, Canada) and one industrial hemp variety ('Anka,' UniSeeds, obtained from Céréla, Saint-Hughes, Québec, Canada) were screened to identify potential polymorphisms in clade V *CsMLOs* that could be associated with increased susceptibility or resistance to PM. Raw sequencing files (Illumina paired-end 125 bp) from these 32 cultivars (BioProject accession: PRJNA738519, SRA accessions: SRR14857079-110) were aligned on the reference CBDRx genome using 'speedseq' v.0.1.2 (Chiang et al., 2015). 'bcftools view' v.1.10.2 (Li, 2011) was used on the raw BAM alignment files to extract the genomic regions corresponding to the two clade V *CsMLO* gene models. Genotypes in these two *CsMLOs* were called for Single Nucleotide Polymorphisms (SNPs) and small insertions and deletions (INDELs) across the 32 cannabis cultivars using 'bcftools mpileup' v.1.10.2 (Li, 2011). Variants with a mapping quality <20 and/or with a read depth >500 \times were filtered out, and allelic frequencies for each variant were extracted using an in-house Python script (v.3.7.3, Python Software Foundation, 2020). Next, SNPGenie v.1.0 (Nelson et al., 2015) was used to estimate the ratio of non-synonymous to synonymous polymorphisms in the two clade V *CsMLO* genes across all 32 cultivars. Based on the output from SNPGenie, genomic positions of all non-synonymous polymorphisms found in the two genes were extracted using an in-house Python script (v.3.7.3, Python Software Foundation, 2020). Visual representations of these

polymorphisms at the gDNA and amino acid levels were prepared using Microsoft® PowerPoint v.16.49 (Microsoft Corporation, Redmond, WA, United States).

RESULTS

CsMLO Gene Identification and Genomic Localization

Through careful manual curation, we were able to identify a total of 14, 17, 19, 18 and 18 distinct *CsMLO* genes in the genomes of CBDRx, Jamaican Lion female, Jamaican Lion male, Purple Kush, and Finola, respectively. Our final manually curated *CsMLO* genes were numbered 1–15 based on chromosomal positions in CBDRx, from chromosome 1 through chromosome X (Table 1, Supplementary Tables 1–5, and Supplementary Files 1–3). *CsMLO* gene numbers and IDs in all four other genomes (Finola, Jamaican Lion female, Jamaican Lion male, Purple Kush) were based on homology relationships, supported by our phylogenetic and orthology analyses (see the section “Materials and Methods” for details). We also physically located each set of *CsMLO* genes on the chromosomes of CBDRx, Finola and Purple Kush (Figure 1). Our results show that eight of the 10 chromosomes harbor evenly spaced *CsMLO* genes, chromosomes 5 and 9 being the only ones not carrying *MLO* genes. We were able to anchor all 14 CBDRx *CsMLO* genes on their respective chromosomes, while two (*CsMLO12-FN*, *CsMLO15-FN*) and six (*CsMLO1-PK_B*, *CsMLO2-PK*, *CsMLO4-PK*, *CsMLO6-PK*, *CsMLO7-PK*, *CsMLO12-PK*) *MLO* genes were located on unanchored contigs in Finola and Purple Kush, respectively (Supplementary Figure 1, Table 1, and Supplementary Tables 1, 4, 5). The CBDRx genome showed an absence of *MLO15*, while this gene was present in all other cannabis genomes. We identified homology and paralogy relationships among all *CsMLO* sequences, which revealed potential duplication patterns in certain genomes. Overall, we detected paralogs for *CsMLO1*, *CsMLO5*, *CsMLO9*, *CsMLO10*, *CsMLO12* and *CsMLO13* in the genomes of Finola, Purple Kush, Jamaican Lion male and Jamaican Lion female (*CsMLO* paralogs were designated with A/B suffixes, see Table 1 and Supplementary Tables 1–5). CBDRx remained the only genome in which we did not identify *CsMLO* paralogs. Most of the *CsMLO* genes had syntenic positions across all three chromosome-level genome assemblies, with the exception of *CsMLO9*, *CsMLO10* and *CsMLO14* (Figure 1). *CsMLO14* was the only *CsMLO* gene identified across all five genomes that had three different locations in the three genomes, i.e., chromosome X in CBDRx, chromosome 1 in Finola and chromosome 4 in Purple Kush. Partial/incomplete genes were also noted in the genomes of Finola (*CsMLO3-FN-B* and *CsMLO10-FN-A*) and Purple Kush (*CsMLO1-PK-B*, *CsMLO2-PK* and *CsMLO5-PK-A*).

CsMLO Gene Structure and Protein Characterization

Manually curated *CsMLO* genes identified in CBDRx ranged in size between 3,567 bp (*CsMLO14-CBDRx*) and

49,009 bp (*CsMLO5-CBDRx*), with an average size of 11,240 bp and a median size of 5,431 bp (Table 1 and Supplementary Tables 1–5). The structural organization of these *CsMLO* genes is depicted in Figure 2. The number of exons varied between 12 and 15, with some of the exons showing signs of fusion in certain genes (Figure 2). The number of amino acid residues in these *CsMLO* gene sequences varied between 509 and 632. Intron size varied considerably, with 11 introns belonging to four different *CsMLO* genes (*CsMLO4*, *CsMLO5*, *CsMLO6*, *CsMLO10*) exhibiting a length greater than 1,000 bp. *CsMLO5* had the longest introns: intron 5 (15.6 kb), intron 9 (7.9 kb), intron 12 (15.9 kb), and intron 14 (6.3 kb) (Figure 2). The *CsMLO* genes that we characterized in the four other genomes (Finola, Jamaican Lion female and male, Purple Kush) were consistently similar to the ones identified in CBDRx in terms of length, intron and exon structural organization and genomic localization (Supplementary Files 1–3). The longest *CsMLO* gene characterized among all our manually curated gene models belonged to Jamaican Lion (female), with a total length of 49,673 nucleotides (*CsMLO5-JL*).

Proteins encoded by all identified *CsMLO* genes comprised seven transmembrane domains (TMs) of similar lengths (Table 1 and Supplementary Tables 1–5). The only exceptions were found in the genomes of Finola and Purple Kush in which we identified a total of four partial *CsMLO* genes, for which encoded proteins harbored less than seven TMs: *CsMLO3-FN-B* (five TMs), *CsMLO10-FN-A* (three TMs), *CsMLO1-PK-B* (four TMs), *CsMLO5-PK-A* (six TMs). Similarly, all identified *CsMLO* proteins were predicted to be localized in the cell membrane by several online prediction servers, such as Plant-mPloc (Chou and Shen, 2010), DeepLoc 1.0 (Almagro Armenteros et al., 2017), and YLoc-HighRes (Briesemeister et al., 2010; Table 1 and Supplementary Tables 1–5). Three partial/incomplete sequences were also predicted to localize elsewhere, such as the chloroplasts and nucleus by Plant-mPloc for *CsMLO1-PK-B*, and the endoplasmic reticulum by DeepLoc 1.0 for *CsMLO5-PK-A* and *CsMLO3-FN-B*. We assumed that these three predictions were unreliable as they were made using incomplete sequences and not present unanimously throughout all prediction servers. No signal peptide (SignalP 5.0, Almagro Armenteros et al., 2019a), mitochondrial transit peptide, chloroplast transit peptide, or thylakoid luminal transit peptide (TargetP 2.0, Almagro Armenteros et al., 2019b) were predicted in any of the *CsMLO* protein sequences (results not shown). The invariable 30 amino acid residues previously described in Elliott et al. (2005) were identified in all *CsMLOs* (Table 1 and Supplementary Tables 1–5). Across all identified *CsMLOs* (excluding the five partial sequences), the amount of conserved amino acids varied between 25 and 30, and 74.1% (60/81) of *CsMLO* sequences possessed all 30 amino acids. In Kusch et al. (2016), a larger dataset of *MLO* proteins was analyzed and thus identified 58 highly conserved amino acids, rather than invariant, showing that substitutions are possible. These 58 amino acid residues were also screened in all *CsMLOs* (Table 1 and Supplementary Tables 1–5). Across all identified *CsMLOs* (excluding the five partial sequences), the amount of conserved amino acids varied between 52 and 58, and only

23.5% (19/81) of CsMLO sequences possessed all 58 amino acids. However, 74.1% (60/81) possessed 57 or more of the conserved amino acids.

Phylogenetic Analysis of CsMLOs

We performed a phylogenetic analysis on the curated cannabis MLO proteins identified among the five genomes (CsMLOs). The dataset was completed with the MLO protein family from *Arabidopsis thaliana* (AtMLOs, Devoto et al., 2003), *Prunus persica* (PpMLOs, Pessina et al., 2014), *Vitis vinifera* (VvMLOs, Feechan et al., 2008), *Hordeum vulgare* and *Zea mays* (HvMLOs and ZmMLOs, Kusch et al., 2016), using the *Chlamydomonas reinhardtii* MLO as the outgroup. Phylogenetic tree construction was performed using the PhyML + SMS algorithm, which confirmed the seven known clades of MLO proteins (Figure 3), with bootstrap values equal or greater than 97% (except for clade I, supported with a bootstrap value of 77%). Clade numbers from I to VII were assigned according to the previous study of Kusch et al. (2016). Previous studies have reported the presence of an eighth clade (e.g., Pessina et al., 2014), clustering with clade VII in other papers. Here, we followed the seven clades nomenclature, but this potential eighth clade would correspond to one of the two clade VII subclades. Indeed, potential subclades were also identified in our study, indicated as separate lines in the inner circle of Figure 3, which were also supported with high values of bootstrap (equal or greater than 83%, with the exception of one clade V subclade, supported with a value of 67%). The same clades and subclades were also found to be supported by high posterior values equal or greater than 77% (indicated in red), following phylogenetic tree construction using MrBayes (also see Supplementary Figure 1). Apart from one clade II subclade that appeared monocot-specific, all subclades depicted here included CsMLOs, as well as PpMLOs and VvMLOs. In each subclade, all CsMLOs clustered together with bootstrap values of 100. Two subclades were found in the phylogenetic clade V, containing all the dicot MLO proteins experimentally shown to be required for PM susceptibility (Acevedo-Garcia et al., 2014). In one of those clade V subclades, three of the cannabis genomes were found to harbor two near-identical genes. All cannabis genomes except CBDRx were found to include one MLO gene grouping with clade IV, which contains all monocot MLO proteins acting as PM susceptibility factors (Figure 3).

Transcriptional Reprogramming of Cannabis MLOs in Response to *Golovinomyces ambrosiae* Infection

We conducted an RNA-seq analysis to look at the transcript abundance of CsMLO genes in leaves of the susceptible cannabis cultivar “Pineapple Express” during infection by *G. ambrosiae* emend. (including *G. spadicus*). Five time points were investigated, corresponding to key stages of the infection: 0 (control), 6 h post-inoculation (hpi) (conidia germination and appressoria formation), 24 hpi (haustoria formation), 3 days post-inoculation (dpi) (secondary hyphae formation), and 8 dpi (secondary haustoria, secondary appressoria and conidiophore/conidia formation). No significant expression was

detected at any time point for CsMLO6, CsMLO8, CsMLO11, and CsMLO15 (in this particular case, reads were aligned to the genome of Jamaican Lion male, as the CBDRx genome is devoid of CsMLO15). While CsMLO3 and CsMLO11 were expressed, no up- or down-regulation was observed under our conditions. Nine genes, namely CsMLO1, CsMLO2, CsMLO4, CsMLO5, CsMLO7, and CsMLO9 were found to be significantly differentially expressed (FDR < 0.05) after inoculation with the pathogen (Figure 4 and Supplementary Figure 2). In the case of clade V genes, CsMLO1 showed a peak of 2.23-fold up-regulation at 6 hpi (FDR = 2.83×10^{-4}), and remained somewhat constant for the remaining of the infection, while the expression of CsMLO4 increased steadily between T0 and 3 dpi, reaching a peak of 2.02-fold up-regulation (when compared to T0) and remained constant at 8 dpi (FDR = 7.87×10^{-4}).

An analysis of the 2 kb upstream region of all CsMLO genes identified through the five cannabis genomes revealed the presence of key regulatory motifs with functions related to environmental/hormonal response (e.g., ABRE, AuxRR-core, ERE, GARE, P-box, TATC-box, TGA-element, GT1-motif, G-box, light response elements), stress and defense response (TC-rich repeats, MBS, ARE, GC-motif, LTR element), developmental regulation (circadian, HD-Zip, CCAAT-box, MSA-like), seed-specific metabolism (O2-site, RY-element) and wound response (WUN-motif). In total, 82 (95%) CsMLO genes had at least one motif related to environmental/hormonal response, while 80 (93%) CsMLO genes had a MYB-related sequence, a motif typically involved in development, metabolism and responses to biotic and abiotic stresses (Dubos et al., 2010; Supplementary Table 6). We found the presence of at least one *cis*-acting element involved in gene overexpression by biotic and abiotic factors (ABRE, CGTCA, TGACG, TCA) in 100% of the 13 CsMLO genes from clade V identified in all five genomes. The analysis of protein domains, their location in the protein and the overall topology of each gene ultimately revealed a consistent pattern among all CsMLO genes.

Subcellular Localization of Clade V CsMLOs

The subcellular localization of clade V MLOs, such as CsMLO1 and CsMLO4, was first analyzed using online tools. As mentioned previously, these analyses predicted that CsMLO1 and CsMLO4 possessed seven transmembrane domains and were localized in the plasma membrane (Table 1 and Supplementary Tables 1–5). To determine the subcellular localization of CsMLO1 and CsMLO4 *in planta*, we constructed two vectors under the control of the CaMV 35S promoter where the coding sequences of CsMLO1 and CsMLO4 were fused to enhanced green fluorescent protein (EGFP) in C-terminal (35S::CsMLO1-EGFP and 35S::CsMLO4-EGFP). The agroinfiltration-based transient gene expression system was used to transform *Nicotiana benthamiana* leaves with each construct. The epidermal cells were observed 3 days after transformation for GFP signal using confocal laser scanning microscopy. The CsMLO1-EGFP fusion protein was observed in the cell periphery as well as throughout

the cell forming networks in a punctuate pattern, while the CsMLO4-EGFP fusion protein was observed solely in the cell periphery in a defined way (Figure 5).

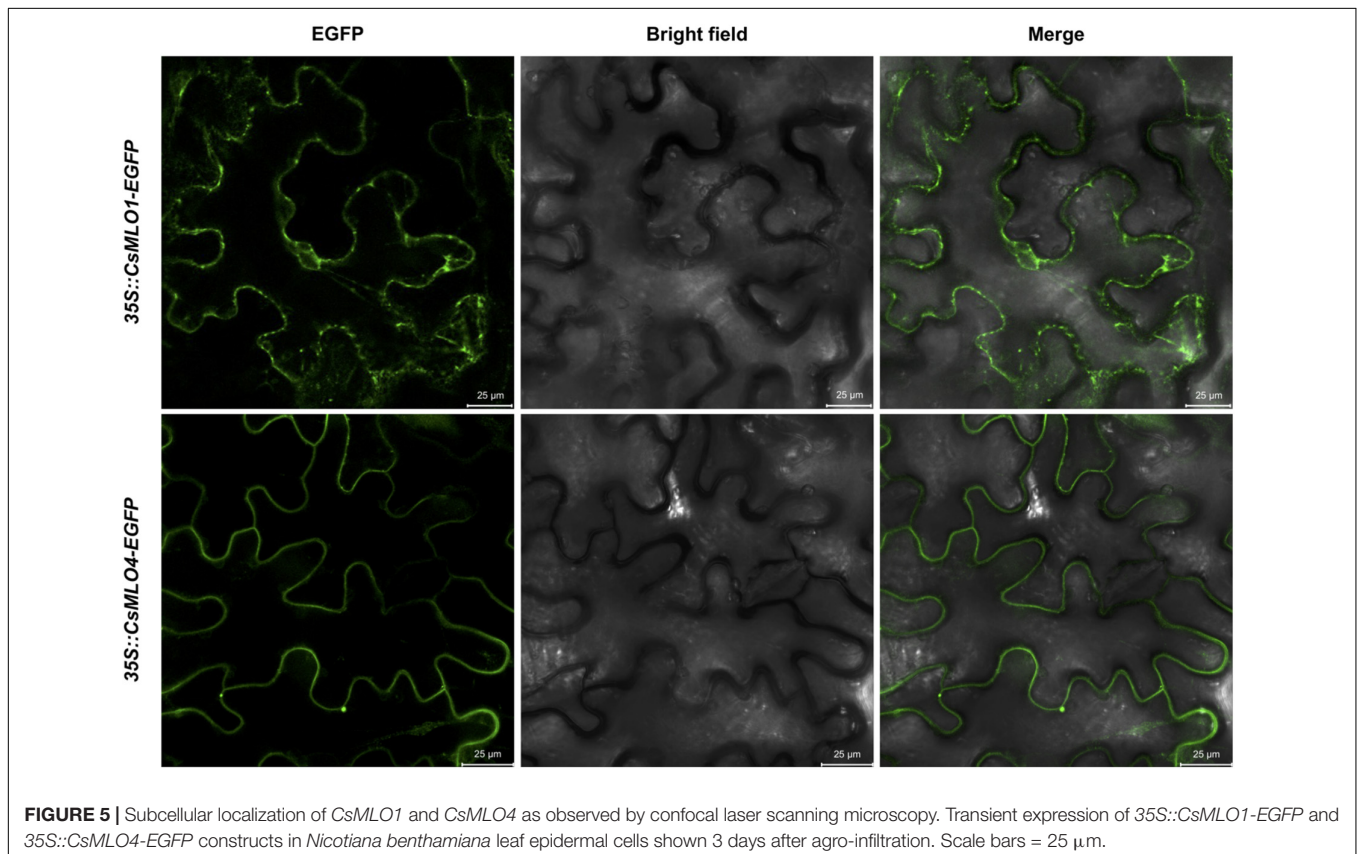
Polymorphism Analysis of Clade V CsMLOs Among 32 Cultivars

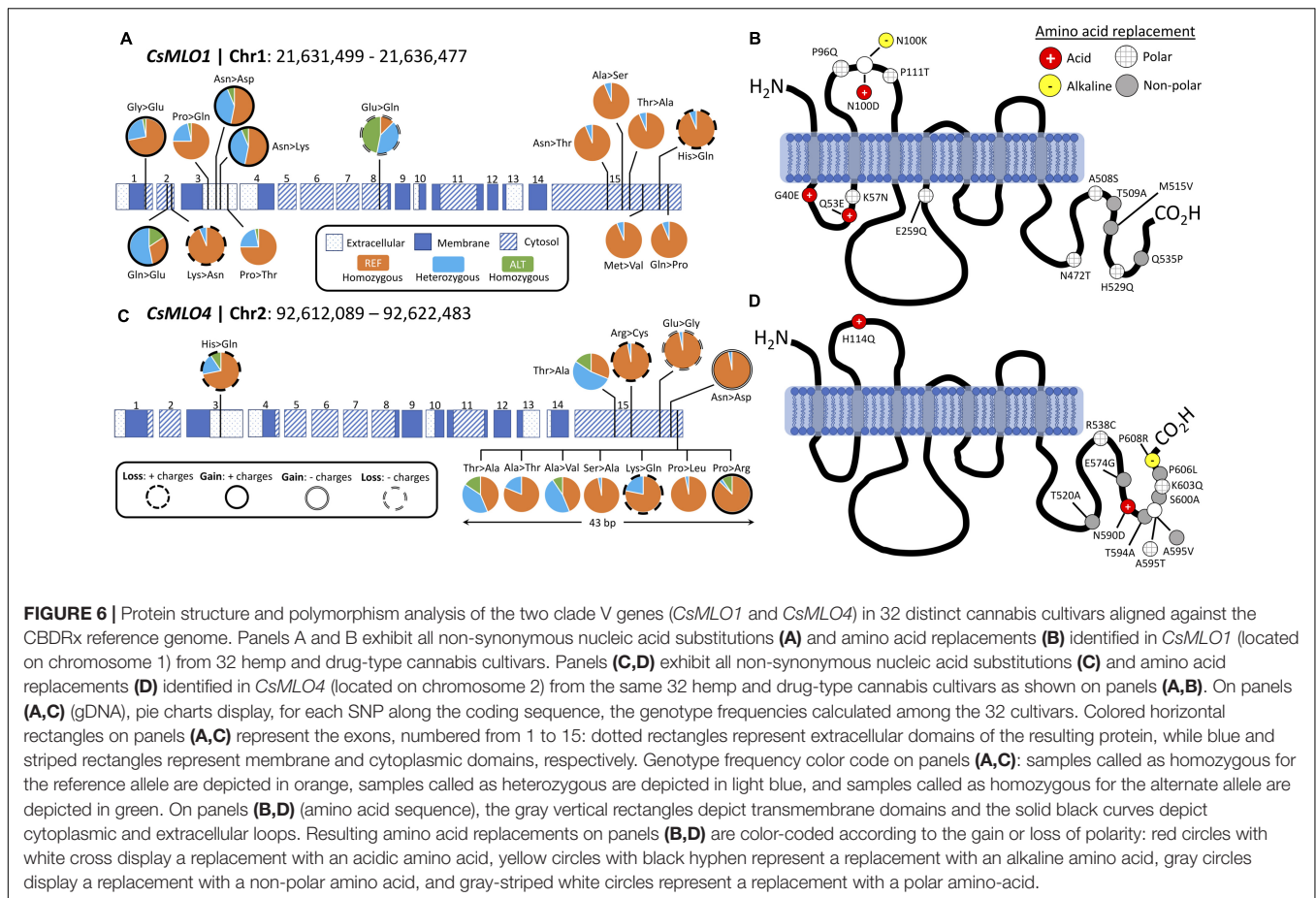
Comparison of 32 distinct cannabis cultivars to the CBDRx reference genome to detect polymorphisms in CsMLO clade V genes revealed a total of 337 (*CsMLO1*, 101 indels and 236 SNPs) and 852 (*CsMLO4*, 154 indels and 698 SNPs) polymorphisms, mainly located in the last cytosolic loop of the protein, near the C-terminus. Among these polymorphic loci, we identified 14 and 12 non-synonymous SNPs for *CsMLO1* and *CsMLO4*, respectively (Supplementary Table 7). The only SNP that was not located near the C- or N-terminus was found in *CsMLO1*, in the second cytoplasmic loop. All other SNPs in *CsMLO1* were distributed in the first cytoplasmic loop (three non-synonymous SNPs), the second extracellular loop (four non-synonymous SNPs) and the last cytoplasmic loop, near the C-terminus (six non-synonymous SNPs, Figure 6). The scenario for *CsMLO4* is slightly different, with a single non-synonymous SNP identified in the second extracellular loop and the remaining 11 SNPs all located in the last cytoplasmic loop, near the C-terminus (Figure 6). About half of these non-synonymous SNPs in both genes induced a change in amino acid charges or polarity, with seven (50%) and five (42%) SNPs having a change in electric charges in *CsMLO1*

and *CsMLO4*, respectively. None of the SNPs identified in either of the two CsMLO clade V genes had an impact on conserved amino acids in the proteins (Elliott et al., 2005; Kusch et al., 2016). Overall, allele frequencies associated with these SNPs in *CsMLO1* and *CsMLO4* showed an even distribution of reference and alternate alleles throughout the 32 cultivars. There was one exception with 'Ultra Sour', which was identified as the only homozygous cultivar for the alternate allele in three non-synonymous SNPs found in exons 1 (G40E) and 3 (P96Q, P111T).

DISCUSSION

Considering the role of specific MLO genes in flowering plants' susceptibility to PM, one of the most prevalent pathogens in indoor cannabis productions (Punja et al., 2019), our primary goal was to structurally and functionally characterize this gene family at a manual-curation level in multiple cannabis genomes. In order to develop mitigation strategies aimed at reducing the deleterious impacts of the pathogen on cannabis production, and in an attempt to better understand other functional roles of CsMLOs, the first step consisted in identifying the exact number and structure of these genes in different genetic backgrounds. Our results first showed that CsMLO numbers are variable across different cannabis types. Second, they showed that two distinct clade V genes were present in all genomes (with paralogs in





certain cultivars) and that these clade V genes possessed *cis*-acting elements typically overexpressed by biotic and abiotic factors. These specific elements in the promoter regions of clade V *CsMLOs* allow them to be responsive to experimental PM infection, which we validated in the context of an infection time point experiment.

CsMLOs and the Importance of Manual Curation

We used 15 *Arabidopsis thaliana* MLO protein sequences to mine the genomes of five cannabis cultivars of four different types (THC-dominant, balanced THC:CBD, CBD-dominant, food-oil hemp), yielding a sum of 86 *CsMLOs* across all genomes (Table 1 and Supplementary Tables 1–5). According to our data, these genes are organized into 15 *CsMLO* homologs in total (Figures 1, 3 and Supplementary Tables 1–5). We were not able to retrieve *CsMLO15* in CBDRx, making this genome devoid of clade IV MLO. Whether this is an artifactual gene loss resulting from the genome assembly and cleaning process or a biological reality in this specific cultivar remains to be verified with additional sequencing data. If this is a biological reality, it could suggest that these genes potentially have redundant and similar functions or are imbricated into functional networks buffered by redundancy (Tully et al., 2014; AbuQamar et al.,

2017). The loss of a gene in the genome of CBDRx in this context could potentially be phenotypically less detrimental. Gene length also varied with a 14-fold difference between the shortest and longest sequence, with two genes (*CsMLO5* and *CsMLO10*) exhibiting multiple unusually long introns (> 10,000 nucleotides, Figure 2). Intron length distribution across the 15 *CsMLOs* showed considerable variability, which explained the significant variations observed in gene length across all *CsMLOs*. Plant introns are typically relatively short and rarely extend beyond 1 kb, making these large *CsMLO* introns up to 10 times longer than typical plant exons (Wu et al., 2013). However, the numbers and positions of exons and introns for the same homologs across the five genomes were highly conserved. We found that cannabis genes tend to have on average five introns (median number of introns across the genome = 4), indicating that *CsMLOs* have three times the number of introns found in a typical gene from the cannabis genome. This could indicate that those introns are likely to play an important functional role and, thus, may be a significant aspect of gene regulation (Seoighe and Korir, 2011). On the other hand, selection against intron size could be counterbalanced in *CsMLOs* by a selective preference for larger introns which correlates with more regulatory elements and a more complex transcriptional control (e.g., in *Vitis vinifera*, Jiang and Goertzen, 2011). Even though the complete *CsMLO* gene catalog could be retrieved in each genome through gene

prediction algorithms combined with targeted BLAST searches, the precise characterization of each gene structure (i.e., start/stop codons, coding sequence, exon–intron boundaries) could not be achieved without multiple efforts of manual curation.

Automated gene structure prediction algorithms are often considered sufficiently reliable to recover the complete genome-wide repertoire of genes of a given sample. However, as repeat content, size, and structural complexity of those genes increase (i.e., numerous small exons delimited by long introns, as observed in *CsMLOs*), errors are increasingly likely to occur and thus impair the accuracy of automated annotations (Guigó et al., 2000; Pilkington et al., 2018). Because 13% of the *CsMLOs* identified here had introns larger than 10 kb (*CsMLO5* and *CsMLO10*), and two additional genes (*CsMLO4* and *CsMLO6*) had introns larger than 1,000 bp, most of these *CsMLOs* were mispredicted by automated gene prediction algorithms (**Supplementary Tables 1–3**). These algorithms typically have 10 kb as the default maximum intron length (e.g., in MAKER2, Holt and Yandell, 2011). In comparison, *Arabidopsis thaliana* genome annotation version 10 indicates that there are 127,854 introns in the nuclear genes, and of these, 99.23% are less than 1,000 bp, while only 16 introns are larger than 5 kb (NCBI accession: GCF_000001735.4, TAIR10.1). Long multi-exon genes having long introns end up fragmented into several shorter “genes” by these programs, thus inflating the actual number of genes within the family. The severity of such errors is influenced by various factors such as the quality of the assembly (Treangen and Salzberg, 2012; Yandell and Ence, 2012) and the availability and quality of extrinsic evidence (e.g., RNA-seq, orthologous sequences). While assembly quality is influenced by genome size and repeat content (Tørresen et al., 2019; Whibley et al., 2021), the disparity in the number of mispredicted genes observed in this study is also likely to be related to differences in sequencing technology (Oxford Nanopore vs. Pacific Biosciences), sequencing depth, algorithms used for *de novo* assembly, and simple base calling accuracy. Overall, these results revealed that cannabis possesses an extensive repertoire of *MLOs* characterized by significant gene size variations across all family members. These results also demonstrate the importance of manual curation when working with automatically generated gene models. Identifying the complete and detailed set of *CsMLOs* for each genome allowed the possibility to assess synteny and evolutionary patterns among *Cannabis sativa* and other plant species.

Evolutionary Dynamics of *CsMLOs*, Gene Duplications and Potential Implications

The vast majority of studies investigating phylogenetic relationships within the *MLO* gene family usually classifies its members in seven clades. A few times, an extra clade has been proposed, or various subclades have been identified, but no consensus has been reached yet. In our study, we classified genes according to the seven clades, and identified some potential subclades presented in **Figure 3**, most of which appear concordant with results from other studies. In two previous studies (Rispaill and Rubiales, 2016; Polanco et al., 2018), clade I is divided in two subclades (Ia and Ib). Subclade Ia would

correspond in our tree to the two subclades respectively including *CsMLO12* and *CsMLO9*, while subclade Ib would correspond to the subclade that includes *CsMLO2*. However, Iovieno et al. (2016) divided clade I in three subclades (Ia, Ib, and Ic). Subclade Ia from Iovieno et al. (2016) appears to correspond to subclade Ib from Polanco et al. (2018), and in our tree is represented by the subclade including *CsMLO2*. Subclades Ib and Ic from Iovieno et al. (2016) would correspond to subclade Ia from Polanco et al. (2018) and would be represented by the subclades including *CsMLO12* and *CsMLO9*, respectively. In Iovieno et al. (2016), clade II is divided into 15 subclades (named from IIa to IIq excluding IIj). Subclades IIa and IIb would be represented in our tree by the subclade including *CsMLO6*; subclades IIc, IId, and IIe would be represented by the monocot-specific subclade; subclades IIg, IIh and IIi would be represented by the subclade including *CsMLO13*; and subclades IIj to IIq would be represented by the subclade including *CsMLO14*. Still in Iovieno et al. (2016), clade III is divided into three subclades (IIIa, IIIb, and IIIc), subclades IIIb and IIIc corresponding in our tree to the subclades including *CsMLO8* and *CsMLO7*, respectively. Subclade IIIa would correspond to the two immediate outlying monocot sequences (*ZmMlo2* and *ZmMlo3*), while our two next outlying monocot sequences (*HvMlo2* and *ZmMlo4*) would correspond to clade VIII (in Iovieno et al., 2016, as there is no consensus on clade VIII). Clade IV has also been divided into two subclades (IVa and IVb) which are grouped together in our tree, subclade IVa from their study simply corresponding to monocot sequences and subclade IVb corresponding to dicot sequences. Clade V has been divided into three subclades (Va, Vb, and Vc), for which subclade Va corresponds in our tree to the subclade including *CsMLO1*, while subclade Vb would correspond to the two outlying sequences from the other subclade from our tree (*PpMLO3* and *VvMLO3*), and subclade Vc would correspond to the remaining of this subclade that includes *CsMLO4*. Clade VI is not divided into subclades in this paper, while we clearly identified two subclades in our tree, one including *CsMLO5* and one including *CsMLO3*. Clade VII has been divided into two subclades in Iovieno et al. (2016), and the same division can be found in our tree, with Iovieno’s subclade VIIa corresponding to the subclade that includes *CsMLO11*, while subclade VIIb would correspond to the subclade including *CsMLO10*. In Zheng et al. (2016), a different clade VIII had been defined, which would correspond to Iovieno et al. (2016) subclade VIIa, while Zheng et al. (2016) clade VII would correspond to Iovieno et al. (2016) subclade VIIb. In most studies, this eighth clade has been merged within clade VII, which is also the case here. As described above, a distinct clade VIII was also defined as a monocot-specific clade in Iovieno et al. (2016), making the use of an 8-clades system confusing. In our opinion, the clade VIII described by Iovieno et al. (2016) could be considered as a subclade of clade III, represented in our tree by the most “diverged” sequences in this clade, *HvMlo2* and *ZmMlo4*.

Manual curation of *CsMLOs* across the five studied genomes revealed with exactitude their respective genomic localization, showing an overall conserved syntenic pattern, except for two genomes, Finola and Purple Kush, which exhibited distinct chromosome localizations for specific *CsMLO* orthologs, as

compared to the rest of the genomes. The number of *CsMLOs* per genome, ranging from 14 to 18, was comparable to other plant genomes, such as *Arabidopsis thaliana* (15, Chen et al., 2006), *Vitis vinifera* (14, Feechan et al., 2008), *Cucumis sativus* (13, Zhou et al., 2013), *Solanum lycopersicum* (15, Chen et al., 2014), *Hordeum vulgare* (11, Kusch et al., 2016), *Medicago truncatula* (14, Rispaill and Rubiales, 2016), *Cicer arietinum* (13, Rispaill and Rubiales, 2016), *Lupinus angustifolius* (15, Rispaill and Rubiales, 2016), *Arachis* spp. (14, Rispaill and Rubiales, 2016), *Cajanus cajan* (20, Rispaill and Rubiales, 2016), *Phaseolus vulgaris* (19, Rispaill and Rubiales, 2016) and *Vigna radiata* (18, Rispaill and Rubiales, 2016). The genomes of Finola and Purple Kush, however, exhibited certain anomalies. Firstly, we found a greater proportion of genes located on unanchored contigs in Finola (11%) and Purple Kush (33%) as compared to CBDRx (0%). Secondly, two paralog pairs, one in Finola (*CsMLO10-FN-A* and *CsMLO10-FN-B*) and one in Purple Kush (*CsMLO9-PK-A* and *CsMLO9-PK-B*), had one of the two paralogs located on a different chromosome (**Figure 1**).

The genome-wide distribution of *CsMLOs* described here did not suggest the involvement of tandem duplications as a predominant mechanism of emergence for *MLOs* in cannabis, as suggested in other taxa (e.g., Liu and Zhu, 2008; Pessina et al., 2014; Rispaill and Rubiales, 2016). Indeed, recent bioinformatic analyses suggested that segmental and tandem duplications were a widespread mechanism for the expansion of the *MLO* gene family in diverse plant species, spanning from algae to dicots (Shi et al., 2020). For example, clear evidence of tandem duplication events have been detected in *P. vulgaris* and *V. radiata*, and in *M. domestica*, respectively (Pessina et al., 2014; Rispaill and Rubiales, 2016). We did not find evidence that tandem duplications were widespread in *CsMLOs*, as most of the genes were evenly spread out in the genomes, with multiple other unrelated genes in between these *CsMLOs*. Some of the *CsMLOs* were located physically close to one another and they were genetically related, but not similar enough (<85%) to be considered as tandem duplicates. In this case, segmental duplication appears to be a more likely mechanism of emergence for *CsMLOs*, although we did not specifically search for segmental duplications in the present study. In total, four (4.7%) *CsMLOs* across the five genomes were located in tandem duplications (**Figure 1**). These four putative tandem gene duplications were located in Finola only, on chromosome X (*CsMLO13-FN-A* and *CsMLO13-FN-B*) and on chromosome 2 (*CsMLO3-FN-A* and *CsMLO3-FN-B*). Other *CsMLOs* that could represent potential tandem gene duplications in Jamaican Lion male (*CsMLO1-JLm*, *CsMLO12-JLm*, *CsMLO13-JLm*) and female (*CsMLO1-JL*, *CsMLO10-JL*) all had the two paralogs located on different contigs that were, on average, longer than 2 Mb each. These different contigs containing two *CsMLO* paralogs typically harbored large sequences (>10 kb) of high homology (>95% similarity), which could suggest that they are either the result of segmental duplications, or that they represent two copies of highly polymorphic loci (Fan et al., 2008; Lallemand et al., 2020). We did not find evidence of tandem duplications in the genome of Purple Kush, indicating that this genome is potentially more fragmented than the others. The duplication of *CsMLO13*

(clade II) on chromosome X is of potential interest as it was duplicated in the male genomes only (Finola and Jamaican Lion male, **Supplementary Tables 3, 5**). *MLO* clade II genes originally evolved in ancient seed-producing plants, suggesting that genes from this clade could have sex-related functions (Feechan et al., 2008; Jiwan et al., 2013; Zhou et al., 2013). On the other hand, this male-specific duplication could represent a technical artifact caused by the fact that the sequence of *CsMLO13* on chromosome Y was concatenated with its homologous version on chromosome X, thus producing a false tandem duplication. To our knowledge, no studies have documented in detail the structure and evolutionary relationships between orthologous *MLOs* in a group of genomes from the same species. Plus, information on sex-specific distribution of *MLOs* in comparable heterogametic sex plant systems is scarce, which makes the interpretation of this finding difficult. This could ultimately indicate that clade II *MLOs* may not be solely related to seed production or development, as suggested in other systems (Kusch et al., 2016). Recently, a clade II *MLO* was shown to be dictating PM susceptibility in mungbean (Yundaeng et al., 2020). However, apart from this example, only clade V *MLOs* have been shown to be involved in this trait in dicots. In this study, one cannabis clade V *MLO*, *CsMLO1*, appears to be duplicated in three different genomes (three out of the four THC-producing cultivars). If not an assembly artifact, the presence of such an additional copy of a clade V *MLO* would make it tedious to obtain complete immunity to PM. In other plants such as *Arabidopsis*, inactivation of all clade V *MLOs* is required to achieve complete immunity, even though these genes unevenly contribute to susceptibility, with *AtMLO2* playing a major role (Consonni et al., 2006). The presence of a single copy of *CsMLO1* in the industrial hemp variety Finola and the CBD-dominant CBDRx (which ancestry has been suggested to be 11% hemp, Grassa et al., 2021) could suggest that hemp is *de facto* less susceptible or that attaining this target in hemp could be easier to achieve. On the other hand, both Jamaican Lion cultivars investigated here have been described as being highly resistant to PM (McKernan et al., 2020), even though both have this extra, apparently functional, *CsMLO1* copy. In this particular case, resistance to PM is likely due to other genetic factors than a loss-of-susceptibility that would have been obtained through deletion/mutation of clade V *MLOs* (see below).

Upregulation of Clade V *MLOs* Triggered by Powdery Mildew Infection

The transcriptomic response of *Cannabis sativa* to PM performed here revealed that clade V *CsMLOs* are rapidly triggered, at least 6 h post-inoculation, upon infection by the pathogen (**Figure 4**). Our results revealed that the two clade V *CsMLO* genes responded with different activation patterns, potentially suggesting specific roles. As described above, all three clade V *MLO* genes need to be inactivated in order to achieve complete immunity against PM in *Arabidopsis* (Consonni et al., 2006). However, in other plants, not all members of clade V are S-genes, and the inactivation of only a subset of the clade V *MLO* genes is required (Bai et al., 2008; Pavan et al., 2011). In those cases, the

precise identification of the exact genes involved in susceptibility needs to rely on additional criteria. A shared element of all *MLO* genes involved in PM susceptibility is that they respond to fungal penetration, showing significant upregulation as soon as 6 h after inoculation (Piffanelli et al., 2002; Bai et al., 2008; Zheng et al., 2013), and candidate genes can thus be identified based on increased expression (Feechan et al., 2008; Pessina et al., 2014). Indeed, being an *S*-gene, the expression of *MLO* is necessary for the successful invasion of PM (Freialdenhoven et al., 1996; Lyngkjær et al., 2000; Zellerhoff et al., 2010). For instance, in watermelon, upregulation was only observed for one clade V *MLO* (out of five), *CIMLO12*, and only at time points corresponding to nine and 24 h after inoculation with *Podosphaera xanthii*, making it the prime candidate dictating PM susceptibility (Iovieno et al., 2015). In apple, three genes (including two out of the four clade V *MLOs*, *MdMLO11* and *MdMLO19*) were found to be significantly up-regulated after inoculation with PM, reaching about 2-fold compared to non-inoculated plants (Pessina et al., 2014). Similarly, three out of four grapevine clade V *MLOs* (*VvMLO3*, *VvMLO4*, and *VvMLO17*) were induced during infection by *Erysiphe necator* (Feechan et al., 2008). Interestingly, in both apple and grapevine, a clade VII *MLO* was also found to respond to PM, but the significance of this finding is unclear. In our case, both clade V *CsMLOs* (*CsMLO1* and *CsMLO4*) appeared responsive to PM infection, showing a 2-fold upregulation (FDR < 0.05). Even though our experimental design was limited, both clade V *CsMLOs* were induced following inoculation with PM, while it was not the case for other *CsMLOs*. Validating the expression of clade V *CsMLOs* using a different approach and/or more importantly using different cultivars would be important to confirm our findings. It is interesting to note that analyzing similar time points (5 and 8 days after inoculation) from a similar RNA-seq experiment (BioProject PRJNA634569), using the same method as the one used in this study, showed a significantly higher expression of both *CsMLO1* and *CsMLO4* at both time points, when compared with mock-inoculated controls (*t*-tests, *P* < 0.05). This was further supported by the presence of *cis*-acting elements involved in gene overexpression by biotic and abiotic stresses in all of the clade V *CsMLOs* characterized in this study. These combined results suggest that cannabis is in a similar situation to that observed in *A. thaliana*, where all clade V genes could be involved in susceptibility.

As of now, *VrMLO12* in mungbean (*Vigna radiata*), which clusters with clade II genes, is the sole report of an *MLO* gene outside of clade V being clearly involved in PM susceptibility in dicots (Yundaeng et al., 2020). In rice, a clade II *MLO* (*OsMLO3*) was also found to have an expression pattern similar to clade V *MLOs* from *Arabidopsis* (Nguyen et al., 2016) and could partially restore PM susceptibility in barley mutants, suggesting an involvement in plant defense (Elliott et al., 2002). According to our results, clade II *CsMLOs* should not be considered candidates, as none showed an induction following inoculation with *G. ambrosiae* emend. (including *G. spadicus*). However, other *CsMLOs* outside of clade V were found to be differentially expressed, a situation not different from previous findings in apple (Pessina et al., 2014), grapevine (Feechan et al., 2008), and

tomato (Zheng et al., 2016). Interestingly, while only one out of four clade V *MLO* showed pathogen-dependent upregulation in tomato (*SIMLO1*), there is some overlap between non-clade V genes that respond to PM in tomato and cannabis. The pathogen-triggered response is, however, not always in the same direction. In tomato, three clade I *MLOs* (*SIMLO10*, *SIMLO13* and *SIMLO14*) were induced following an infection challenge with *Oidium neolyopersici*. While the cannabis orthologs of the last two (*CsMLO9* and *CsMLO2*, respectively) were also differentially expressed after inoculation with PM, their expression levels rather decreased over time (Supplementary Figure 2). Similarly, the expression of the tomato clade III *SIMLO4* significantly increased during infection, while that of its cannabis ortholog *CsMLO7* decreased (Supplementary Figure 2). The sole tomato clade VI gene, *SIMLO16*, was also found to be induced (Zheng et al., 2016). While its direct cannabis ortholog *CsMLO3* was not found to be up-regulated, it was the case for the other cannabis clade VI gene, *CsMLO5*, for which no ortholog exists in tomato or *Arabidopsis*. This particular gene was the only *CsMLO* to be induced following a challenge with *G. ambrosiae* emend. (including *G. spadicus*) outside of clade V.

Following our attempt to further characterize the two clade V *CsMLOs* and determine their subcellular localization *in planta* using confocal laser scanning microscopy, we observed that *CsMLO4* was located in the plasma membrane while *CsMLO1* was located in endomembrane-associated compartments, including the plasma membrane, the endoplasmic reticulum and the Golgi apparatus (Figure 5). In plants, the presence of a reticulate and network-looking pattern and bright spots, as observed for *CsMLO1*, are typically associated with the endoplasmic reticulum and Golgi stacks, respectively (Bassham et al., 2008). A time-series performed using confocal microscopy demonstrated that the *CsMLO1*-EGFP fusion protein was extremely dynamic compared to the *CsMLO4*-EGFP fusion protein (results not shown), supporting its implication in intracellular trafficking through the endomembrane system. Many studies have demonstrated that *MLOs* are localized in the plasma membrane (Devoto et al., 1999; Kim and Hwang, 2012; Nie et al., 2015), supporting our observations with regards to *CsMLO4*. Other studies have indicated that *MLOs* are associated with the plasma membrane and/or other endomembrane compartments, such as the endoplasmic reticulum and the Golgi apparatus (Chen et al., 2009; Jones and Kessler, 2017; Qin et al., 2019) and thus supporting our observations with regards to *CsMLO1*. To further support our findings and to determine precise subcellular localization of *CsMLOs*, subcellular fractionation studies as well as fluorescence colocalization with specific organelle markers should be performed.

Genetic Variation Within Clade V *CsMLOs*: A Quest Toward Durable Resistance

Several natural or induced loss-of-function mutations have been identified in *MLO* genes that reduce susceptibility to PM (Büschges et al., 1997; Consonni et al., 2006; Bai et al., 2008; Pavan et al., 2011; Wang et al., 2014). Barley *HvMlo1*

is probably the most studied *MLO* gene, and most mutations leading to loss-of-function (and thus, resistance to PM) in this gene appear to cluster in the second and third cytoplasmic loops (Reinstädler et al., 2010). The functional importance of those loops is still unclear, but evidence in other plants also points toward this particular region (Fujimura et al., 2016; Acevedo-Garcia et al., 2017). Outside of this region, the integrity of transmembrane domains, as well as certain invariant cysteine and proline residues, are critical for the function and accumulation of *MLO* proteins (Elliott et al., 2005; Reinstädler et al., 2010). Examination of polymorphisms among 32 cannabis cultivars (including one hemp variety) identified only a single SNP in the second cytoplasmic loop of *CsMLO1*. No mutations were identified among those invariant cysteine and proline residues, and no mutations were found either in any of the strictly conserved residues or within transmembrane domains. This suggests that loss-of-function mutations in *CsMLOs* could be rare or non-existent among commercial cultivars, complicating future breeding efforts. The potential lack of diversity among the cultivars included in our analysis might also have impeded our ability to find causative loss-of-function mutations, which could be a scenario likely generalizable to a significant part of the cannabis production industry. Nevertheless, it is possible that mutations identified outside of those previously identified regions could inactivate *CsMLOs*. Three mutations causing amino acid replacements were identified in the first cytoplasmic loop of *CsMLO1*, and five mutations (four in *CsMLO1* and one in *CsMLO4*) were identified in the first extracellular loop. The proximal part of the C-terminus of *MLO* proteins contains a binding site for the cytoplasmic calcium sensor, calmodulin (Kim et al., 2002a,b). In barley *MLO*, binding of calmodulin to this domain appears to be required for full susceptibility. Unfortunately, even though a high number of polymorphisms were identified in the cytoplasmic C-terminus of both *CsMLO1* (six mutations) and *CsMLO4* (10 mutations), those mutations are not found within the calmodulin-binding domain but rather at the distal end of the C-terminus. This might not be surprising, as this region is intrinsically disordered (Kusch et al., 2016). Intrinsically disordered regions, i.e., regions lacking stable secondary structures, usually exhibit greater amounts of non-synonymous mutations and other types of polymorphisms because of the lack of structural constraints (Nilsson et al., 2011; Kusch et al., 2016). Another study conducted on clade V *MLOs* had also revealed that both the first extracellular loop and the C-terminus were under strong positive selection (Iovieno et al., 2015), which seems in agreement with our observations.

The fact that potential loss-of-function mutations in clade V *MLOs* were not identified among the 32 investigated genomes suggests that complete resistance to PM might be hard to find among existing commercial cultivars. Nevertheless, such mutations might be present at a very low frequency, especially in “wild” populations or in landraces that have infrequently been used in breeding programs. For example, natural *mlo* alleles exist in barley but have only been found in landraces from Ethiopia and Yemen (Reinstädler et al., 2010). Similarly, the natural loss-of-function mutations in pea and tomato originated from wild accessions (Bai et al., 2008; Humphry et al., 2011).

This should reinforce the importance of preserving cannabis wild populations and encourages efforts to establish germplasm repositories. However, *MLO*-based resistance to PM being a recessive trait, and assuming that both *CsMLO1* and *CsMLO4* are involved, this would mean that loss-of-function mutations would need to be bred as homozygous recessive for both genes into elite plants (not considering that multiple copies of a given *CsMLO* might exist, as suggested here for *CsMLO1*). In the absence of natural mutants, or to accelerate the implementation of *mlo*-based resistance in breeding programs, induced mutagenesis and genome editing might be interesting alternatives. While such approaches have not been optimized for cannabis, loss-of-function mutations have been obtained through those in other crops, such as barley (Reinstädler et al., 2010), wheat (Wang et al., 2014; Acevedo-Garcia et al., 2017), or tomato (Nekrasov et al., 2017).

There might also be other routes to combating PM than *MLO*. In crops where gene-for-gene interactions exist with PM (e.g., in cereals), a series of functional alleles confer complete resistance against distinct sets of PM isolates (Bourras et al., 2019). Since a similar situation has been observed between hop (*Humulus lupulus*, the closest relative of cannabis) and *Podosphaera macularis* (Henning et al., 2011), it is likely that such gene-for-gene interactions exist between cannabis and PM. Indeed, the first gene conferring complete resistance to an isolate of PM has recently been identified in cannabis (Mihalyov and Garfinkel, 2021). In grapevine, resistance is usually considered polygenic, and there appears to be a diverse range of responses to invasion by *Erysiphe necator*, from penetration resistance to the induction of plant cell death (Feechan et al., 2011). Because cannabis has been excluded from scientific research for decades, very little is known about the biology of PM infection (i.e., genetic diversity, histology, host range), and no data on resistance levels among cultivars has been published as of now. Still, a recent study observed higher copy numbers of a thaumatin-like protein in several cannabis cultivars reported to be resistant to PM, and similar correlations were identified with copy number variations of endochitinases and *CsMLOs* (McKernan et al., 2020). Thaumatin-like proteins and endochitinases might thus represent additional targets potentially involved in quantitative resistance to PM.

CONCLUSION

Genome-wide characterization of *CsMLOs* performed in this study indicated that genes from clade V, which are immediately triggered upon infection by PM, are likely involved as early actors in the fungal infection process by the plant. Polymorphism data generated for clade V *CsMLOs* in multiple commercial cultivars suggested that loss-of-function mutations, required to achieve a resistance phenotype, are rare events and potentially challenging to assemble, especially when considering their recessive nature and the genetic redundancy of multiple clade V *CsMLOs*. Preserving a diversified collection of feral and, ideally, landrace cannabis genetic backgrounds while establishing coherent germplasm repositories could represent efficient strategies to compose with this complex biological reality. This will allow the

creation of novel and valuable knowledge on the fundamental biology supporting the interaction between PM and cannabis, ultimately leading to more sustainable horticultural and agro-industrial practices.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: National Center for Biotechnology Information (NCBI) BioProject database under accession number PRJNA738505, SRA accessions: SRR14839036-SRR14839050. Python and Shell BASH scripts (total of 3) are provided as a compressed archive file (**Supplementary File 4**).

AUTHOR CONTRIBUTIONS

NP, FOH, and DLJ contributed to the conception and design of the study. NP performed the experiments. NP, FOH, and DLJ analyzed the data and drafted the manuscript. DLJ contributed reagents, equipment, and funds. All authors revised the manuscript, read, and approved the submitted version.

FUNDING

TRICHUM (Translating Research into Innovation for Cannabis Health at Université de Moncton) is supported by grants from Genome Canada (Genome Atlantic NB-RP3), the Atlantic Canada Opportunity Agency (project 212090), and the New Brunswick Innovation Foundation (RIF2018-036), Mitacs and Organigram.

ACKNOWLEDGMENTS

We greatly thank François Sormany and Alex J. Cull for their help generating the RNA-seq and genomic datasets, respectively, Hugo Germain and his team for sharing their expertise with confocal microscopy, and Organigram for their continued support and for providing biomaterials.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.729261/full#supplementary-material>

REFERENCES

- Abbasi, W. A., Asif, A., Andleeb, S., and Minhas, F. U. A. A. (2017). CaMELS: *in silico* prediction of calmodulin binding proteins and their binding sites. *Proteins* 85, 1724–1740. doi: 10.1002/prot.25330
- Ablazov, A., and Tombuloglu, H. (2016). Genome-wide identification of the mildew resistance locus O (*MLO*) gene family in novel cereal model species

Supplementary Figure 1 | Phylogenetic relationships of CsMLOs based on Bayesian inference analysis. Phylogenetic tree of manually curated CsMLO proteins (bold) with MLO proteins from selected species (*Arabidopsis thaliana*, *Prunus persica*, *Vitis vinifera*, *Hordeum vulgare*, and *Zea mays*). *Chlamydomonas reinhardtii* was used as an outgroup. Phylogenetic relationships were estimated using the MrBayes tool implemented on NGPhylogeny.fr, using default parameters. The seven defined clades are indicated, as well as potential subclades identified in this study (inner circles). Number on a node indicates the posterior probabilities of major clades and subclades. MLOs with one asterisk (*) have been experimentally demonstrated to be required for PM susceptibility (Büsches et al., 1997; Feechan et al., 2008; Wan et al., 2020), while MLOs with two asterisks (**) have been identified as main probable candidates for PM susceptibility (Pessina et al., 2016).

Supplementary Figure 2 | Transcriptomic response of CsMLO genes from all clades, except from clade V, following inoculations with powdery mildew. CsMLO genes from clades I, II, III, VI, and VII are displayed in clade-specific plots with different colors depicting the different clades: red (clade I), gray (clade II), yellow (clade III), orange (clade VI), and purple (clade VII). Gene expression is displayed on the y-axis as the average logarithmic value of the counts per million [$\log_2(\text{CPM})$] at each time point (displayed on the x-axis, $n =$ per time point). Time points: no infection/control (T0), 6 h post-inoculation (6H), 24 h post-inoculation (1D), 3 days post-inoculation (3D), and 8 days post-inoculation (8D). Error bars at each time point represent the standard deviation (SD).

Supplementary Table 1 | Members of the CsMLO gene family as predicted and manually curated in the CBDRx genome.

Supplementary Table 2 | Members of the CsMLO gene family as predicted and manually curated in the Jamaican Lion (female) genome.

Supplementary Table 3 | Members of the CsMLO gene family as predicted and manually curated in the Jamaican Lion (male) genome.

Supplementary Table 4 | Members of the CsMLO gene family as predicted and manually curated in the Purple Kush genome.

Supplementary Table 5 | Members of the CsMLO gene family as predicted and manually curated in the Finola genome.

Supplementary Table 6 | Plant *cis*-acting regulatory elements identified in CsMLOs across five distinct cultivars.

Supplementary Table 7 | Non-synonymous SNPs identified in CsMLO1 and CsMLO4 of 32 cannabis cultivars with their respective amino acid change in the resulting protein.

Supplementary File 1 | Nucleotide alignment of the genomic sequences from all 86 CsMLOs. Exons and introns are separated by multiple gaps in the alignment, as is the case for the insertion in the first exon of CsMLO5-JLm.

Supplementary File 2 | Nucleotide alignment of the coding sequences from all 86 CsMLOs.

Supplementary File 3 | Amino acid alignment of the protein sequences from all 86 CsMLOs.

Supplementary File 4 | Python and Bash scripts used to parse SNP data and RNA-seq results (archive file).

- Brachypodium distachyon*. *Eur. J. Plant Pathol.* 145, 239–253. doi: 10.1007/s10658-015-0833-2
- AbuQamar, S., Moustafa, K., and Tran, L. S. (2017). Mechanisms and strategies of plant defense against *Botrytis cinerea*. *Crit. Rev. Biotechnol.* 37, 262–274. doi: 10.1080/07388551.2016.1271767
- Acevedo-Garcia, J., Kusch, S., and Panstruga, R. (2014). Magical mystery tour: MLO proteins in plant immunity and beyond. *New Phytol.* 204, 273–281. doi: 10.1111/nph.12889

- Acevedo-Garcia, J., Spencer, D., Thieron, H., Reinstädler, A., Hammond-Kosack, K., Phillips, A. L., et al. (2017). *mlo*-based powdery mildew resistance in hexaploid bread wheat generated by a non-transgenic TILLING approach. *Plant Biotechnol. J.* 15, 367–378. doi: 10.1111/pbi.12631
- Almagro Armenteros, J. J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., et al. (2019a). Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* 2:e201900429. doi: 10.26508/lsa.201900429
- Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* 33, 3387–3395. doi: 10.1093/bioinformatics/btx431
- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019b). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z
- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Appiano, M., Pavan, S., Catalano, D., Zheng, Z., Bracuto, V., Lotti, C., et al. (2015). Identification of candidate *MLO* powdery mildew susceptibility genes in cultivated Solanaceae and functional characterization of tobacco *NtMLO1*. *Transgenic Res.* 24, 847–858. doi: 10.1007/s11248-015-9878-4
- Bai, Y., Pavan, S., Zheng, Z., Zappel, N. F., Reinstädler, A., Lotti, C., et al. (2008). Naturally occurring broad-spectrum powdery mildew resistance in a Central American tomato accession is caused by loss of *Mlo* function. *Mol. Plant Microbe Interact.* 21, 30–39. doi: 10.1094/MPMI-21-1-0030
- Bassham, D. C., Brandizzi, F., Otegui, M. S., and Sanderfoot, A. A. (2008). The secretory system of *Arabidopsis*. *Arabidopsis Book* 6:e0116. doi: 10.1199/tab.0116
- Bidzinski, P., Noir, S., Shahi, S., Reinstädler, A., Gratkowska, D. M., and Panstruga, R. (2014). Physiological characterization and genetic modifiers of aberrant root thigmomorphogenesis in mutants of *Arabidopsis thaliana* *MILDEW LOCUS O* genes. *Plant Cell Environ.* 37, 2738–2753. doi: 10.1111/pce.12353
- Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, D344–D354. doi: 10.1093/nar/gkaa977
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bourras, S., Kunz, L., Xue, M., Praz, C. R., Müller, M. C., Kälin, C., et al. (2019). The *AvrPm3-Pm3* effector-NLR interactions control both race-specific resistance and host-specificity of cereal mildews on wheat. *Nat. Commun.* 10:2292. doi: 10.1038/s41467-019-10274-1
- Braun, U., and Cook, R. T. A. (2012). *Taxonomic Manual of the Erysiphales* (Powdery Mildews). CBS Biodiversity Series No. 11. Utrecht: CBS-KNAW Fungal Biodiversity Centre.
- Braun, U., Cook, R. T. A., Inman, A. J., and Shin, H. D. (2002). “The taxonomy of the powdery mildew fungi,” in *The Powdery Mildews: A Comprehensive Treatise*, eds R. R. Bélanger, W. R. Bushnell, A. J. Dik, and T. L. W. Carver (Saint Paul, MN: APS Press), 13–55.
- Briesemeister, S., Rahnenführer, J., and Kohlbacher, O. (2010). YLoc-an interpretable web server for predicting subcellular localization. *Nucleic Acids Res.* 38, 497–502. doi: 10.1093/nar/gkq477
- Büsches, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., et al. (1997). The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell* 88, 695–705. doi: 10.1016/S0092-8674(00)81912-1
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Chen, Y., Wang, Y., and Zhang, H. (2014). Genome-wide analysis of the mildew resistance locus o (*MLO*) gene family in tomato (*Solanum lycopersicum* L.). *Plant Omics* 7, 87–93.
- Chen, Z., Hartmann, H. A., Wu, M.-J., Friedman, E. J., Chen, J.-G., Pulley, M., et al. (2006). Expression analysis of the *AtMLO* gene family encoding plant-specific seven-transmembrane domain proteins. *Plant Mol. Biol.* 60, 583–597. doi: 10.1007/s11103-005-5082-x
- Chen, Z., Noir, S., Kwaaitaal, M., Hartmann, H. A., Wu, M. J., Mudgil, Y., et al. (2009). Two seven-transmembrane domain MILDEW RESISTANCE LOCUS O proteins cofunction in *Arabidopsis* root thigmomorphogenesis. *Plant Cell* 21, 1972–1991. doi: 10.1105/tpc.108.062653
- Chiang, C., Layer, R. M., Faust, G. G., Lindberg, M. R., David, B., Garrison, E. P., et al. (2015). SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat. Methods* 12, 966–968. doi: 10.1038/nmeth.3505
- Chou, K. C., and Shen, H. B. (2010). Plant-mPLOC: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One* 5:e11335. doi: 10.1371/journal.pone.0011335
- Consonni, C., Humphry, M. E., Hartmann, H. A., Livaja, M., Durner, J., Westphal, L., et al. (2006). Conserved requirement for a plant host cell protein in powdery mildew pathogenesis. *Nat. Genet.* 38, 716–720. doi: 10.1038/ng1806
- Dangl, J. L., and Jones, J. D. G. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–833. doi: 10.1038/35081161
- Debode, F., Janssen, E., and Berben, G. (2013). Development of 10 new screening PCR assays for GMO detection targeting promoters (pFMV, pNOS, pSuAra, pTA29, pUbi, pRice actin) and terminators (t35S, tE9, tOCS, tg7). *Eur. Food Res. Technol.* 236, 659–669. doi: 10.1007/s00217-013-1921-1
- Deshmukh, R., Singh, V. K., and Singh, B. D. (2014). Comparative phylogenetic analysis of genome-wide *Mlo* gene family members from *Glycine max* and *Arabidopsis thaliana*. *Mol. Genet. Genomics* 289, 345–359. doi: 10.1007/s00438-014-0811-y
- Devoto, A., Hartmann, H. A., Piffanelli, P., Elliott, C., Simmons, C., Taramino, G., et al. (2003). Molecular phylogeny and evolution of the plant-specific seven-transmembrane *MLO* family. *J. Mol. Evol.* 56, 77–88. doi: 10.1007/s00239-002-2382-5
- Devoto, A., Piffanelli, P., Nilsson, I. M., Wallin, E., Panstruga, R., von Heijne, G., et al. (1999). Topology, subcellular localization, and sequence diversity of the *Mlo* family in plants. *J. Biol. Chem.* 274, 34993–35004. doi: 10.1074/jbc.274.49.34993
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Dobson, L., Reményi, I., and Tusnády, G. E. (2015). CCTOP: a consensus constrained topology prediction web server. *Nucleic Acids Res.* 43, W408–W412. doi: 10.1093/nar/gkv451
- Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 15, 573–581. doi: 10.1016/j.tplants.2010.06.005
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Elliott, C., Müller, J., Miklis, M., Bhat, R. A., Schulze-Lefert, P., and Panstruga, R. (2005). Conserved extracellular cysteine residues and cytoplasmic loop-loop interplay are required for functionality of the heptahelical *MLO* protein. *Biochem. J.* 385, 243–254. doi: 10.1042/BJ20040993
- Elliott, C., Zhou, F., Spielmeier, W., Panstruga, R., and Schulze-Lefert, P. (2002). Functional conservation of wheat and rice *Mlo* orthologs in defense modulation to the powdery mildew fungus. *Mol. Plant Microbe Interact.* 15, 1069–1077. doi: 10.1094/MPMI.2002.15.10.1069
- Fan, C., Chen, Y., and Long, M. (2008). Recurrent tandem gene duplication gave rise to functionally divergent genes in *Drosophila*. *Mol. Biol. Evol.* 25, 1451–1458. doi: 10.1093/molbev/msn089
- Farinas, C., and Peduto Hand, F. (2020). First report of *Golovinomyces spadicus* causing powdery mildew on industrial hemp (*Cannabis sativa*) in Ohio. *Plant Dis.* 104:2727. doi: 10.1094/PDIS-01-20-0198-PDN
- Feechan, A., Jermakow, A. M., Torregrosa, L., Panstruga, R., and Dry, I. B. (2008). Identification of grapevine *MLO* gene candidates involved in susceptibility to powdery mildew. *Funct. Plant Biol.* 35, 1255–1266. doi: 10.1071/FP08173
- Feechan, A., Kabbara, S., and Dry, I. B. (2011). Mechanisms of powdery mildew resistance in the Vitaceae family. *Mol. Plant Pathol.* 12, 263–274. doi: 10.1111/j.1364-3703.2010.00668.x
- Freialdenhoven, A., Peterhänsel, C., Kurth, J., Kreuzaler, F., and Schulze-Lefert, P. (1996). Identification of genes required for the function of non-race-specific *mlo* resistance to powdery mildew in barley. *Plant Cell* 8, 5–14. doi: 10.1105/tpc.8.1.5

- Freisleben, R., and Lein, A. (1942). Über die Auffindung einer mehltreueresistenten Mutante nach Röntgenbestrahlung einer anfälligen reinen Linie von Sommergerste. *Naturwissenschaften* 30:608. doi: 10.1007/BF01488231
- Fujimura, T., Sato, S., Tajima, T., and Arai, M. (2016). Powdery mildew resistance in the Japanese domestic tobacco cultivar Kokubu is associated with aberrant splicing of *MLO* orthologues. *Plant Pathol.* 65, 1358–1365. doi: 10.1111/ppa.12498
- Glawe, D. A. (2008). The powdery mildews: a review of the world's most familiar (yet poorly known) plant pathogens. *Annu. Rev. Phytopathol.* 46, 27–51. doi: 10.1146/annurev.phyto.46.081407.104740
- Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Timothy Motley, S., et al. (2021). A new *Cannabis* genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytol.* 230, 1665–1679. doi: 10.1111/nph.17243
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812. doi: 10.1093/bioinformatics/btu393
- Guigó, R., Agarwal, P., Abril, J. F., Burset, M., and Fickett, J. W. (2000). An assessment of gene prediction accuracy in large DNA sequences. *Genome Res.* 10, 1631–1642. doi: 10.1101/gr.122800
- Hartley, J. L., Temple, G. F., and Brasch, M. A. (2000). DNA cloning using *in vitro* site-specific recombination. *Genome Res.* 10, 1788–1795. doi: 10.1101/gr.143000
- Health Canada (2020). *Licensed Cultivators, Processors and Sellers of Cannabis Under the Cannabis Act*. Available online at: <https://www.canada.ca/en/health-canada/services/drugs-medication/cannabis/industry-licenses-applicants/licensed-cultivators-processors-sellers.html> (accessed December 1, 2020)
- Henning, J. A., Townsend, M. S., Gent, D. H., Bassil, N., Matthews, P., Buck, E., et al. (2011). QTL mapping of powdery mildew susceptibility in hop (*Humulus lupulus* L.). *Euphytica* 180, 411–420. doi: 10.1007/s10681-011-0403-4
- Hilbert, M., Novero, M., Rovenich, H., Mari, S., Grimm, C., Bonfante, P., et al. (2020). *MLO* differentially regulates barley root colonization by beneficial endophytic and mycorrhizal fungi. *Front. Plant Sci.* 10:1678. doi: 10.3389/fpls.2019.01678
- Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. doi: 10.1186/1471-2105-12-491
- Huelsensbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754
- Humphry, M., Reinstädler, A., Ivanov, S., Bisseling, T., and Panstruga, R. (2011). Durable broad-spectrum powdery mildew resistance in pea *er1* plants is conferred by natural loss-of-function mutations in *PsMLO1*. *Mol. Plant Pathol.* 12, 866–878. doi: 10.1111/j.1364-3703.2011.00718.x
- Ingvarsdén, C. R., Massange-Sánchez, J. A., Borum, F., Uauy, C., and Gregersen, P. L. (2019). Development of *mlo*-based resistance in tetraploid wheat against wheat powdery mildew. *Theor. Appl. Genet.* 132, 3009–3022. doi: 10.1007/s00122-019-03402-4
- Iovieno, P., Andolfo, G., Schiavulli, A., Catalano, D., Ricciardi, L., Frusciantè, L., et al. (2015). Structure, evolution and functional inference on the *Mildew Locus O* (*MLO*) gene family in three cultivated Cucurbitaceae spp. *BMC Genomics* 16:1112. doi: 10.1186/s12864-015-2325-3
- Iovieno, P., Bracuto, V., Pavan, S., Lotti, C., Ricciardi, L., and Andolfo, G. (2016). Identification and functional inference on the *Mlo*-family in Viridiplantae. *J. Plant Pathol.* 98, 587–594. doi: 10.4454/JPP.V98I3.027
- Jacott, C. N., Charpentier, M., Murray, J. D., and Ridout, C. J. (2020). *Mildew Locus O* facilitates colonization by arbuscular mycorrhizal fungi in angiosperms. *New Phytol.* 227, 343–351. doi: 10.1111/nph.16465
- Jiang, K., and Goertzen, L. R. (2011). Spliceosomal intron size expansion in domesticated grapevine (*Vitis vinifera*). *BMC Res. Notes* 4:52. doi: 10.1186/1756-0500-4-52
- Jiwan, D., Roalson, E. H., Main, D., and Dhingra, A. (2013). Antisense expression of peach mildew resistance locus *O* (*PpMlo1*) gene confers cross-species resistance to powdery mildew in *Fragaria x ananassa*. *Transgenic Res.* 22, 1119–1131. doi: 10.1007/s11248-013-9715-6
- Jones, D. S., and Kessler, S. A. (2017). Cell type-dependent localization of *MLO* proteins. *Plant Signal. Behav.* 12:e1393135. doi: 10.1080/15592324.2017.1393135
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Jørgensen, I. H. (1992). Discovery, characterization and exploitation of *Mlo* powdery mildew resistance in barley. *Euphytica* 63, 141–152. doi: 10.1007/BF00023919
- Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/s1360-1385(02)02251-3
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kaufmann, H., Qiu, X., Wehmeyer, J., and Debener, T. (2012). Isolation, molecular characterization, and mapping of four rose *MLO* orthologs. *Front. Plant Sci.* 3:244. doi: 10.3389/fpls.2012.00244
- Kessler, S. A., Shimosato-Asano, H., Keinath, N. F., Wuest, S. E., Ingram, G., Panstruga, R., et al. (2010). Conserved molecular components for pollen tube reception and fungal invasion. *Science* 330, 968–971. doi: 10.1126/science.1195211
- Kim, D. S., and Hwang, B. K. (2012). The pepper *MLO* gene, *CaMLO2*, is involved in the susceptibility cell-death response and bacterial and oomycete proliferation. *Plant J.* 72, 843–855. doi: 10.1111/tpj.12003
- Kim, M. C., Lee, S. H., Kim, J. K., Chun, H. J., Choi, M. S., Chung, W. S., et al. (2002a). *Mlo*, a modulator of plant defense and cell death, is a novel calmodulin-binding protein: isolation and characterization of a rice *Mlo* homologue. *J. Biol. Chem.* 277, 19304–19314. doi: 10.1074/jbc.M108478200
- Kim, M. C., Panstruga, R., Elliott, C., Möller, J., Devoto, A., Yoon, H. W., et al. (2002b). Calmodulin interacts with *MLO* protein to regulate defence against mildew in barley. *Nature* 416, 447–450. doi: 10.1038/416447a
- Konishi, S., Sasakuma, T., and Sasanuma, T. (2010). Identification of novel *Mlo* family members in wheat and their genetic characterization. *Genes Genet. Syst.* 85, 167–175. doi: 10.1266/ggs.85.167
- Kusch, S., Pesch, L., and Panstruga, R. (2016). Comprehensive phylogenetic analysis sheds light on the diversity and origin of the *MLO* family of integral membrane proteins. *Genome Biol. Evol.* 8, 878–895. doi: 10.1093/gbe/evw036
- Lallemand, T., Leduc, M., Landès, C., Rizzon, C., and Lerat, E. (2020). An overview of duplicated gene detection methods: why the duplication mechanism has to be accounted for in their choice. *Genes* 11:1046. doi: 10.3390/genes11091046
- Lavery, K. U., Stout, J. M., Sullivan, M. J., Shah, H., Gill, N., Holbrook, L., et al. (2019). A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the *THC/CBD acid synthase* loci. *Genome Res.* 29, 146–156. doi: 10.1101/gr.242594.118
- Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15:R29. doi: 10.1186/gb-2014-15-2-r29
- Lefort, V., Longueville, J. E., and Gascuel, O. (2017). SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149
- Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., et al. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res.* 47, W260–W265. doi: 10.1093/nar/gkz303
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van De Peer, Y., et al. (2002). PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325
- Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Liu, Q., and Zhu, H. (2008). Molecular evolution of the *MLO* gene family in *Oryza sativa* and their functional divergence. *Gene* 409, 1–10. doi: 10.1016/j.gene.2007.10.031

- Lyngkjær, M. F., and Carver, T. L. W. (2000). Conditioning of cellular defence responses to powdery mildew in cereal leaves by prior attack. *Mol. Plant Pathol.* 1, 41–49. doi: 10.1046/j.1364-3703.2000.00006.x
- Lyngkjær, M. F., Newton, A. C., Atzema, J. L., and Baker, S. J. (2000). The barley *mlo*-gene: an important powdery mildew resistance source. *Agronomie* 20, 745–756. doi: 10.1051/agro:2000173
- McKernan, K. J., Helbert, Y., Kane, L. T., Ebling, H., Zhang, L., Liu, B., et al. (2018). Cryptocurrencies and zero mode wave guides: an unclouded path to a more contiguous *Cannabis sativa* L. genome assembly. *OSF [Preprint]* Available online at: <https://osf.io/7d968/> (accessed June 3, 2021).
- McKernan, K., Helbert, Y., Kane, L., Ebling, H., Zhang, L., Liu, B., et al. (2020). Sequence and annotation of 42 *Cannabis* genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes. *Biorxiv [Preprint]* doi: 10.1101/2020.01.03.894428
- Mihalyov, P. D., and Garfinkel, A. R. (2021). Discovery and genetic mapping of PM1, a powdery mildew resistance gene in *Cannabis sativa* L. *Front. Agron.* 3:720215. doi: 10.3389/fagro.2021.720215
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonhammer, E. L. L., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913
- Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D., and Kamoun, S. (2017). Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci. Rep.* 7:482. doi: 10.1038/s41598-017-00578-x
- Nelson, C. W., Moncla, L. H., and Hughes, A. L. (2015). SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics* 31, 3709–3711. doi: 10.1093/bioinformatics/btv449
- Nguyen, V. N. T., Vo, K. T. X., Park, H., Jeon, J. S., and Jung, K. H. (2016). A systematic view of the *MLO* family in rice suggests their novel roles in morphological development, diurnal responses, the light-signaling pathway, and various stress responses. *Front. Plant Sci.* 7:1413. doi: 10.3389/fpls.2016.01413
- Nie, J., Wang, Y., He, H., Guo, C., Zhu, W., Pan, J., et al. (2015). Loss-of-function mutations in *CsMLO1* confer durable powdery mildew resistance in cucumber (*Cucumis sativus* L.). *Front. Plant Sci.* 6:1155. doi: 10.3389/fpls.2015.01155
- Nilsson, J., Grahn, M., and Wright, A. P. H. (2011). Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. *Genome Biol.* 12:R65. doi: 10.1186/gb-2011-12-7-r65
- Pavan, S., Schiavulli, A., Appiano, M., Marcotrigiano, A. R., Cillo, F., Visser, R. G. F., et al. (2011). Pea powdery mildew *er1* resistance is associated to loss-of-function mutations at a *MLO* homologous locus. *Theor. Appl. Genet.* 123, 1425–1431. doi: 10.1007/s00122-011-1677-6
- Pépin, N., Punja, Z. K., and Joly, D. L. (2018). Occurrence of powdery mildew caused by *Golovinomyces cichoracearum sensu lato* on *Cannabis sativa* in Canada. *Plant Dis.* 102:2644. doi: 10.1094/PDIS-04-18-0586-PDN
- Pessina, S., Lenzi, L., Perazzolli, M., Campa, M., Dalla Costa, L., Urso, S., et al. (2016). Knockdown of *MLO* genes reduces susceptibility to powdery mildew in grapevine. *Hortic. Res.* 3:16016. doi: 10.1038/hortres.2016.16
- Pessina, S., Pavan, S., Catalano, D., Gallotta, A., Visser, R. G. F., Bai, Y., et al. (2014). Characterization of the *MLO* gene family in Rosaceae and gene expression analysis in *Malus domestica*. *BMC Genomics* 15:618. doi: 10.1186/1471-2164-15-618
- Piffanelli, P., Zhou, F., Casais, C., Orme, J., Jarosch, B., Schaffrath, U., et al. (2002). The barley *MLO* modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. *Plant Physiol.* 129, 1076–1085. doi: 10.1104/pp.010954
- Pilkington, S. M., Crowhurst, R., Hilario, E., Nardoza, S., Fraser, L., Peng, Y., et al. (2018). A manually annotated *Actinidia chinensis* var. *chinensis* (kiwifruit) genome highlights the challenges associated with draft genomes and gene prediction in plants. *BMC Genomics* 19:257. doi: 10.1186/s12864-018-4656-3
- Polanco, C., Sáenz de Miera, L. E., Bett, K., and de la Vega, M. P. (2018). A genome-wide identification and comparative analysis of the lentil *MLO* genes. *PLoS One* 13:e0194945. doi: 10.1371/journal.pone.0194945
- Punja, Z. K. (2021). Emerging diseases of *Cannabis sativa* and sustainable management. *Pest Manag. Sci.* 77, 3857–3870. doi: 10.1002/ps.6307
- Punja, Z. K., Collyer, D., Scott, C., Lung, S., Holmes, J., and Sutton, D. (2019). Pathogens and molds affecting production and quality of *Cannabis sativa* L. *Front. Plant Sci.* 10:1120. doi: 10.3389/fpls.2019.01120
- Python Software Foundation (2020). *Python Language Reference, Version 3.7.3*. Wilmington, DE: Python Software Foundation.
- Qin, B., Wang, M., He, H.-X., Xiao, H.-X., Zhang, Y., and Wang, L.-F. (2019). Identification and characterization of a potential candidate *MLO* gene conferring susceptibility to powdery mildew in rubber tree. *Phytopathology* 109, 1236–1245. doi: 10.1094/PHYTO-05-18-0171-R
- Reinstädler, A., Müller, J., Czembor, J. H., Piffanelli, P., and Panstruga, R. (2010). Novel induced *mlo* mutant alleles in combination with site-directed mutagenesis reveal functionally important domains in the heptahelical barley *MLO* protein. *BMC Plant Biol.* 10:31. doi: 10.1186/1471-2229-10-31
- Rispail, N., and Rubiales, D. (2016). Genome-wide identification and comparison of legume *MLO* gene family. *Sci. Rep.* 6:32673. doi: 10.1038/srep32673
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- RStudio Team (2020). *RStudio: Integrated Development for R*. Boston, MA: RStudio, PBC. Available online at: <http://www.rstudio.com/>
- Seoighe, C., and Korir, P. K. (2011). Evidence for intron length conservation in a set of mammalian genes associated with embryonic development. *BMC Bioinformatics* 12(Suppl. 9):S16. doi: 10.1186/1471-2105-12-S9-S16
- Shen, Q., Zhao, J., Du, C., Xiang, Y., Cao, J., and Qin, X. (2012). Genome-scale identification of *MLO* domain-containing genes in soybean (*Glycine max* L. Merr.). *Genes Genet. Syst.* 87, 89–98. doi: 10.1266/ggs.87.89
- Shi, J., Wan, H., Zai, W., Xiong, Z., and Wu, W. (2020). Phylogenetic relationship of plant *MLO* genes and transcriptional response of *MLO* genes to *Ralstonia solanacearum* in tomato. *Genes* 11:487. doi: 10.3390/genes11050487
- Singh, V. K., Singh, A. K., Chand, R., and Singh, B. D. (2012). Genome wide analysis of disease resistance *MLO* gene family in sorghum [*Sorghum bicolor* L. Moench]. *J. Plant Genomics* 2, 18–27.
- Small, E. (2015). Evolution and classification of *Cannabis sativa* (Marijuana, Hemp) in relation to human utilization. *Bot. Rev.* 81, 189–294. doi: 10.1007/s12229-015-9157-3
- Statistics Canada (2021). *T20Table 20-10-0008-01 Retail Trade Sales by Province and Territory (x 1,000)*. Ottawa, ON: Statistics Canada.
- Szarka, D., Tymon, L., Amsden, B., Dixon, E., Judy, J., and Gauthier, N. (2019). First report of powdery mildew caused by *Golovinomyces spadiceus* on industrial hemp (*Cannabis sativa*) in Kentucky. *Plant Dis.* 103:1773. doi: 10.1094/PDIS-01-19-0049-PDN
- Thompson, G. R., Tuscano, J. M., Dennis, M., Singapuri, A., Libertini, S., Gaudino, R., et al. (2017). A microbiome assessment of medical marijuana. *Clin. Microbiol. Infect.* 23, 269–270. doi: 10.1016/j.cmi.2016.12.001
- Torresen, O. K., Star, B., Mier, P., Andrade-Navarro, M. A., Bateman, A., Jarnot, P., et al. (2019). Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Res.* 47, 10994–11006. doi: 10.1093/nar/gkz841
- Treangen, T. J., and Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13, 36–46. doi: 10.1038/nrg3117
- Tully, J. P., Hill, A. E., Ahmed, H. M. R., Whitley, R., Skjellum, A., and Mukhtar, M. S. (2014). Expression-based network biology identifies immune-related functional modules involved in plant defense. *BMC Genomics* 15:421. doi: 10.1186/1471-2164-15-421
- Van Rossum, G., and Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- van Schie, C. C. N., and Takken, F. L. W. (2014). Susceptibility genes 101: how to be a good host. *Annu. Rev. Phytopathol.* 52, 551–581. doi: 10.1146/annurev-phyto-102313-045854
- Vielba-Fernández, A., Polonio, Á., Ruiz-Jiménez, L., de Vicente, A., Pérez-García, A., and Fernández-Ortuño, D. (2020). Fungicide resistance in powdery mildew fungi. *Microorganisms* 8, 1–34. doi: 10.3390/microorganisms8091431
- Wan, D. Y., Guo, Y., Cheng, Y., Hu, Y., Xiao, S., Wang, Y., et al. (2020). CRISPR/Cas9-mediated mutagenesis of *VvMLO3* results in enhanced resistance to powdery mildew in grapevine (*Vitis vinifera*). *Hortic. Res.* 7:116. doi: 10.1038/s41438-020-0339-8
- Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers

- heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi: 10.1038/nbt.2969
- Weldon, W. A., Ullrich, M. R., Smart, L. B., Smart, C. D., and Gadoury, D. M. (2020). Cross-infectivity of powdery mildew isolates originating from hemp (*Cannabis sativa*) and Japanese Hop (*Humulus japonicus*) in New York. *Plant Heal. Prog.* 21, 47–53. doi: 10.1094/PHP-09-19-0067-RS
- Whibley, A., Kelley, J. L., and Narum, S. R. (2021). The changing face of genome assemblies: guidance on achieving high-quality reference genomes. *Mol. Ecol. Resour.* 21, 641–652. doi: 10.1111/1755-0998.13312
- Win, K. T., Zhang, C., and Lee, S. (2018). Genome-wide identification and description of MLO family genes in pumpkin (*Cucurbita maxima* Duch.). *Hortic. Environ. Biotechnol.* 59, 397–410. doi: 10.1007/s13580-018-0036-9
- Wiseman, M. S., Bates, T., Garfinkel, A., Ocamb, C. M., and Gent, D. H. (2021). First report of powdery mildew caused by *Golovinomyces ambrosiae* on *Cannabis sativa* in Oregon. *Plant Dis.* doi: 10.1094/PDIS-11-20-2455-PDN
- Wu, J. Y., Xiao, J. F., Wang, L. P., Zhong, J., Yin, H. Y., Wu, S. X., et al. (2013). Systematic analysis of intron size and abundance parameters in diverse lineages. *Sci. China Life Sci.* 56, 968–974. doi: 10.1007/s11427-013-4540-y
- Yandell, M., and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* 13, 329–342. doi: 10.1038/nrg3174
- Yi, J., An, S., and An, G. (2014). *OsMLO12*, encoding seven transmembrane proteins, is involved with pollen hydration in rice. *Plant Reprod.* 27, 169–180. doi: 10.1007/s00497-014-0249-8
- Yundaeng, C., Somta, P., Chen, J., Yuan, X., Chankaew, S., Srinives, P., et al. (2020). Candidate gene mapping reveals *VrMLO12* (MLO Clade II) is associated with powdery mildew resistance in mungbean (*Vigna radiata* [L.] Wilczek). *Plant Sci.* 298:110594. doi: 10.1016/j.plantsci.2020.110594
- Zellerhoff, N., Himmelbach, A., Dong, W., Bieri, S., Schaffrath, U., and Schweizer, P. (2010). Nonhost resistance of barley to different fungal pathogens is associated with largely distinct, quantitative transcriptional responses. *Plant Physiol.* 152, 2053–2066. doi: 10.1104/pp.109.151829
- Zheng, Z., Appiano, M., Pavan, S., Bracuto, V., Ricciardi, L., Visser, R. G. F., et al. (2016). Genome-wide study of the tomato *SIMLO* gene family and its functional characterization in response to the powdery mildew fungus *Oidium neolycopersici*. *Front. Plant Sci.* 7:380. doi: 10.3389/fpls.2016.00380
- Zheng, Z., Nonomura, T., Appiano, M., Pavan, S., Matsuda, Y., Toyoda, H., et al. (2013). Loss of function in *Mlo* orthologs reduces susceptibility of pepper and tomato to powdery mildew disease caused by *Leveillula taurica*. *PLoS One* 8:e70723. doi: 10.1371/journal.pone.0070723
- Zhou, S. J., Jing, Z., and Shi, J. L. (2013). Genome-wide identification, characterization, and expression analysis of the MLO gene family in *Cucumis sativus*. *Genet. Mol. Res.* 12, 6565–6578. doi: 10.4238/2013.December.11.8
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2021 Pépin, Hebert and Joly. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.