1 **Genome-wide characterization of the *MLO* gene family in *Cannabis***

2 ***sativa* reveals two genes as strong candidates for powdery mildew**

3 **susceptibility**

4

5 **Noémi Pépin[1][†], Francois Olivier Hebert[1,2][†], and David L. Joly[1][*]**

6

7 [1] Centre d'innovation et de recherche sur le cannabis (CIRC), Université de Moncton,

8 Département de biologie, Moncton (NB), Canada.

9 [2] Institut National des Cannabinoïdes (INC), Montréal (QC), Canada.

10

11 [*] **Correspondence:**

12 David L. Joly

13 david.joly@umoncton.ca

14 † These authors have contributed equally to this work.

15

16 **Keywords: MLO, Cannabis, Powdery mildew, Fungal disease, Susceptibility genes, Plant-**

17 **pathogen interactions**

18

19 Words: 10,990

20 Figures: 8 (6 Figures + 2 Supplementary Figures)

21 Tables: 8 (1 Table + 7 Supplementary Tables)

22 **Abstract**

23 *Cannabis sativa* is increasingly being grown around the world for medicinal, industrial, and

24 recreational purposes. As in all cultivated plants, cannabis is exposed to a wide range of pathogens,

25 including powdery mildew (PM). This fungal disease stresses cannabis plants and reduces flower

26 bud quality, resulting in significant economic losses for licensed producers. The *Mildew Locus O*

27 (*MLO*) gene family encodes plant-specific proteins distributed among conserved clades, of which

28 clades IV and V are known to be involved in susceptibility to PM in monocots and dicots,

29 respectively. In several studies, the inactivation of those genes resulted in durable resistance to the

30 disease. In this study, we identified and characterized the *MLO* gene family members in five

31 different cannabis genomes. Fifteen *Cannabis sativa MLO* (*CsMLO*) genes were manually curated

32 in cannabis, with numbers varying between 14, 17, 19, 18, and 18 for CBDRx, Jamaican Lion

33 female, Jamaican Lion male, Purple Kush, and Finola, respectively (when considering paralogs

34 and incomplete genes). Further analysis of the *CsMLO* genes and their deduced protein sequences

35 revealed that many characteristics of the gene family, such as the presence of 7 transmembrane

36 domains, the MLO functional domain, and particular amino acid positions, were present and well

37 conserved. Phylogenetic analysis of the MLO protein sequences from all five cannabis genomes

38 and other plant species indicated seven distinct clades (I through VII), as reported in other crops.

39 Expression analysis revealed that the *CsMLOs* from clade V, *CsMLO1* and *CsMLO4*, were

40 significantly upregulated following *Golovinomyces ambrosiae* infection, providing preliminary

41 evidence that they could be involved in PM susceptibility. Finally, the examination of variation

42 within *CsMLO1* and *CsMLO4* in 32 cannabis cultivars revealed several amino acid changes, which

43 could affect their function. Altogether, cannabis *MLO* genes were identified and characterized,

44 among which candidates potentially involved in PM susceptibility were noted. The results of this

45 study will lay the foundation for further investigations, such as the functional characterization of

46 clade V MLOs as well as the potential impact of the amino acid changes reported. Those will be

47 useful for breeding purposes in order to develop resistant cultivars.

48

49

50

1   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, 2021). 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; Weldon et al., 2020; Farinas and Peduto Hand, 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

82 risk products (e.g., potassium bicarbonate, knotweed extract), and physical methods (e.g., de-

83 leafing, irradiation) (Punja, 2021). Nonetheless, some of these methods increase production costs,

84 are labor-intensive, and necessitate further research. Therefore, identifying sources of genetic

85 resistance to PM in cannabis and ultimately breeding or developing resistant cultivars offers the

86 most effective and sustainable approach to controlling PM.

87

88 A common strategy used in resistance breeding relies on the exploitation of resistance

89 genes in plants, which encodes for cytoplasmic receptors such as nucleotide-binding leucine-rich

90 repeat proteins or surface receptors such as receptor-like kinases and receptor-like proteins. These

91 immune receptors can detect specific proteins or molecules produced by the pathogen and

92 subsequently induce plant defense responses (Dangl and Jones, 2001). While useful, most

93 resistance genes confer race-specific resistance and are therefore frequently overcome by the

94 emergence of a pathogen's new virulent race within a few years. An alternative approach in

95 resistance breeding is to exploit susceptibility genes (*S-genes*) in plants. *S-genes* are defined as

96 genes that facilitate infection and support compatibility for a pathogen (van Schie and Takken,

97 2014). The alteration of such genes can limit the pathogen's ability to infect the plant and therefore

98 provide a durable type of resistance (van Schie and Takken, 2014). Such PM resistance was

99 initially observed in an X-ray irradiated barley (*Hordeum vulgare*) population in the 1940s

100 (Freisleben and Lein, 1942). It was discovered later that the immunity was attributable to a mutated

101 *S-gene* named *Mildew Locus O* (*MLO*), which was recessively inherited. Complete resistance to

102 all known isolates of PM, caused by *Blumeria graminis* f. sp. *hordei* was conferred in barley by

103 these loss-of-function mutations when present in the homozygous state. This type of resistance has

104 been durable under field conditions and has been used for over forty years in barley breeding

105 programs, without any break in the resistance (Jørgensen, 1992; Büschges et al., 1997; Lyngkjær

106 and Carver, 2000; Piffanelli et al., 2002).

107

108 Since discovering *MLO* genes in barley, many *MLO* homologs have been identified in

109 several plant species, especially in monocots and eudicots, as the PM disease affects angiosperms

110 solely. For instance, *MLO* genes were identified in Rosaceae (roses (Kaufmann et al., 2012), apple,

111 peach, strawberry and apricot (Pessina et al., 2014)), Cucurbitaceae (cucumber (Zhou et al., 2013),

112 melon, watermelon, zucchini (Iovieno et al., 2015) and pumpkin (Win et al., 2018)), Solanaceae

113   (tomato (Bai et al., 2008), pepper (Zheng et al., 2013), tobacco, potato and eggplant (Appiano et

114   al., 2015)), Fabaceae (pea (Humphry et al., 2011; Pavan et al., 2011), soybean (Shen et al., 2012;

115   Deshmukh et al., 2014), barrel medic, chickpea, narrow-leaf lupin, peanut, pigeon pea, common

116   bean, mung bean (Rispail and Rubiales, 2016) and lentil (Polanco et al., 2018)), Brassicaceae (thale

117   cress (Devoto et al., 1999; Devoto et al., 2003)), Vitaceae (grapevine (Feechan et al., 2008)), and

118   Poaceae (rice (Liu and Zhu, 2008), wheat (Konishi et al., 2010), sorghum (Singh et al., 2012),

119   maize (Devoto et al., 2003; Kush et al., 2016), foxtail millet (Kush et al., 2016) and stiff brome

120   (Ablazov and Tombuloglu, 2016)). Furthermore, thorough phylogenetic analyses of land plants

121   revealed that *MLO* genes were not only present in monocots and eudicots but also in basal

122   angiosperms, gymnosperms, lycophytes, and bryophytes (Jiwan et al., 2013; Kusch et al., 2016;

123   Shi et al., 2020). Many MLO-like proteins were also identified in algae and other unicellular

124   eukaryotes, suggesting that *MLO* is an ancient eukaryotic protein (Jiwan et al., 2013; Kusch et al.,

125   2016; Shi et al., 2020).

126

127   The *MLO* gene family is described as a medium-sized plant-specific gene family, with a

128   varying number of members between 7 in wheat to 39 in soybean, depending on the species

129   (Acedevo-Garcia et al., 2014). The resulting MLO proteins are characterized by the presence of

130   seven transmembrane domains integral to the plasma membrane with an extracellular N-terminus

131   and an intracellular C-terminus (Devoto et al., 1999). They are also characterized by the presence

132   of a calmodulin-binding domain in the C-terminal region that is likely implicated in sensing

133   calcium influx and mediating various signaling cascades (Kim et al., 2002a; Kim et al., 2002b).

134   MLO protein sequences identified across land plants also possess several highly conserved amino

135   acids, some of which have been deemed essential for the structure, functionality, and stability of

136   the protein (Devoto et al., 2003; Elliott et al., 2005; Reinstädler et al., 2010; Kusch et al., 2016).

137   Mutations in these residues could affect the accumulation, maturation, and function of the protein

138   and are therefore attractive targets for breeding programs (Elliott et al., 2005; Reinstädler et al.,

139   2010).

140

141   Throughout land plant evolution, the MLO protein family diversified into subfamilies, or

142   clades, which have been demonstrated in several phylogenetic analyses. MLO proteins are usually

143   grouped into seven defined clades (I to VII), among which clades IV and V appear to host MLO

144 proteins associated with PM susceptibility in monocots and dicots, respectively (Acevedo-Garcia

145 et al., 2014; Kusch et al., 2016). It has been documented in many species, such as barley, tomato,

146 and apple, that *MLO* genes from these two clades (IV and V in monocots and dicots, respectively)

147 are up-regulated upon PM infection (Piffanelli et al., 2002; Zheng et al., 2013; Pessina et al., 2014).

148 It has also been demonstrated that the overexpression of these genes results in enhanced

149 susceptibility to the pathogen (Zheng et al., 2013). Furthermore, the inactivation of these genes in

150 many species by gene silencing, genome editing, or TILLING has resulted in increased or

151 complete resistance to PM (Wang et al., 2014; Acevedo-Garcia et al., 2017; Nekrasov et al., 2017;

152 Ingvardsen et al., 2019; Wan et al., 2020). Besides the implication of clade V and IV MLOs in PM

153 susceptibility, recent studies have suggested that *MLO* genes from other clades are implicated in

154 various physiological and developmental processes. For example, it was demonstrated that in

155 *Arabidopsis thaliana*, *AtMLO4* and *AtMLO11* from clade I are involved in root

156 thigmomorphogenesis (Chen et al., 2009; Bidzinski et al., 2014), while *AtMLO7* from clade III is

157 involved in pollen tube reception by the embryo sac (Kessler et al., 2010). In rice (*Oryza sativa*),

158 *OsMLO12* from clade III mediates pollen hydration (Yi et al., 2014). Interestingly, barley

159 *HvMLO1* has been shown to differentially regulate the establishment of mutualistic interactions

160 with the endophyte *Serendipita indica* and the arbuscular mycorrhizal fungus *Funneliformis*

161 *mosseae* (Hilbert et al., 2020). Indeed, another study clearly showed barley *HvMLO1*, wheat

162 *TaMLO1,* and barrel medic *MtMLO8* from clade IV to be involved in the establishment of

163 symbiotic relationships with beneficial mycorrhizal fungi (Jacott et al., 2020). These findings

164 suggest that powdery mildews might have appropriated and exploited these genes as an entryway

165 to successful pathogenic colonization (Jacott et al., 2020). However, in pea, no evidence was found

166 for the implication of *PsMLO1*, a clade V gene, in the establishment of relationships with

167 mycorrhizal and rhizobial symbionts (Humphry et al., 2011).

168

169   Although *MLO* genes have been studied in many monocot and dicot species, they have

170 only been preliminarily studied in cannabis (McKernan et al., 2020). The growing interest in

171 cannabis research has led to the publication of several genomes in recent years, thus providing an

172 opportunity to conduct a comprehensive analysis of the *MLO* gene family in cannabis. In this

173 study, we manually curated and characterized the members of the *MLO* gene family in cannabis

174 from five different available genomes: Purple Kush and Finola (Laverty et al., 2019), CBDRx

175   (Grassa et al., 2021), and Jamaican Lion (female, McKernan et al., 2018; male, McKernan et al.,

176   2020). Through phylogenetic analysis, we identified candidate *MLO* genes likely to be involved

177   in PM susceptibility in cannabis, observed their subcellular localization by confocal microscopy,

178   and monitored their expression profile in cannabis leaves during infection. We also searched for

179   potential naturally occurring resistant mutants by investigating amino acid changes in 32 cultivars.

180   A better understanding of cannabis *MLOs* offers enormous opportunities to breed PM-resistant

181   cultivars and develop new control methods, thereby increasing productivity and yield.

182

183   **2   Materials and methods**

184

185   **2.1   *In silico* identification and manual curation of the cannabis *MLO* genes**

186   Cannabis *MLO* genes in CBDRx were initially identified (also named 'cs10', NCBI accession

187   GCA_900626175.2) using TBLASTN from the BLAST+ suite (Camacho et al., 2009) with

188   *Arabidopsis thaliana* amino acid sequences as queries (*AtMLO1-AtMLO15*, NCBI accession

189   numbers: NP_192169.1,   NP_172598.1,   NP_566879.1,   NP_563882.1,   NP_180923.1,

190   NP_176350.1,   NP_179335.3,   NP_565416.1,   NP_174980.3,   NP_201398.1,   NP_200187.1,

191   NP_565902.1, NP_567697.1, NP_564257.1, NP_181939.1). In parallel, all official gene models

192   were extracted from the NCBI CBDRx annotation report (NCBI Annotation Release 100,

193   https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Cannabis_sativa/100/) with an InterPro

194   (IPR004326) and/or Pfam (PF03094) identification number related to the *MLO* gene family (Blum

195   et al., 2021; Mistry et al., 2021). These two sequence datasets were merged together, and multiple

196   sequence alignments were performed using MUSCLE v.3.8.31 (Edgar, 2004) with the genomic

197   and the amino acid versions of each *MLO* gene model. Only unique sequences were kept, and all

198   *MLO* gene models that were retained but incomplete were further manually curated. The full

199   genomic sequences were aligned and manually compared with their corresponding full-length

200   mRNA transcripts using BLASTN from the BLAST+ suite (Camacho et al., 2009). Each gene was

201   characterized based on total length, chromosome localization, strand, START and STOP positions,

202   as well as number and size of exons and introns (**Figure 1**, **Table 1**). The resulting streamlined

203   and manually curated *MLO* gene models for CBDRx were considered the definitive reference set

204   for this genome.

205

206   The reference set of CBDRx *MLO* genes were used as queries to search for the presence of

207   homologs in four other cannabis genomes (Purple Kush - GCA_000230575.5, Finola -

208   GCA_003417725.2, Jamaican Lion female - GCA_012923435.1, Jamaican Lion male -

209   GCA_013030025.1), using TBLASTN. Manual curation of each set of *MLO* genes was performed

210   in each of these genomes, using the same approach described previously for CBDRx. In the

211   process, several frameshifts (mostly small insertions and deletions) were noted and manually

212   corrected in the coding sequence of multiple *MLOs* in Finola (13 genes) and Purple Kush (7 genes).

213   None of the homologs for these genes showed any frameshift in any other genome. All the

214   frameshifts identified in Finola and Purple Kush were thus examined by comparing their respective

215   coding sequence with available transcriptomic data from the CanSat3 assembly project

216   (http://genome.ccbr.utoronto.ca/downloads.html), using BLASTN. Based on evidence from

217   mRNA sequences, all these frameshifts were manually corrected. All of the manually curated *MLO*

218   genes for all five cannabis genomes were ultimately considered as our reference and final set of

219   *Cannabis sativa MLO* (*CsMLO*) genes (**Table 1, Supplementary Tables S1-S5, Supplementary**

220   **Files S1-S3**). The structure of our final *CsMLO* gene models was compared to their respective

221   genome annotation available on NCBI (CBDRx, Jamaican Lion male and female)

222   (**Supplementary Tables S1, S2 and S3**).

223

224   Chromosomal localization of each manually curated *CsMLO* gene in the chromosome-

225   level genome assemblies, i.e. CBDRx, Finola, and Purple Kush, was displayed on a chromosomal

226   map (**Figure 2**) with the R package RIdeogram (Hao et al., 2020). Chromosome numbers in Finola

227   and Purple Kush were standardized to match the official chromosome numbers in the CBDRx

228   reference genome, using whole genome alignments on the D-Genies platform

229   (http://dgenies.toulouse.inra.fr/). The SVG output from RIdeogram was aesthetically modified and

230   adapted using Microsoft © PowerPoint v.16.49 (Microsoft Corporation, Redmond, Washington,

231   U.S.). This analysis served as a preliminary assessment of synteny for *CsMLO* genes.

232

233   2.2   Gene and protein characterization

234   The protein sequences from the five genomes were analyzed through several online prediction

235   servers in order to identify functional domains (InterProScan, Jones et al., 2014), transmembrane

236   domains (CCTOP, Dobson et al., 2015), subcellular localizations (Plant-mPLoc, Chou and Shen,

237 2010; DeepLoc 1.0, Almagro Armenteros et al., 2017; YLoc-HighRes, Briesemeister et al., 2010),

238 signal peptide (SignalP 5.0, Almagro Armenteros et al., 2019a), calmodulin-binding domains

239 (CaMELS, Abbasi et al., 2017) as well as mitochondrial, chloroplast, and thylakoid luminal transit

240 peptide (TargetP 2.0, Almagro Armenteros et al., 2019b). Conserved amino acids described in

241 Elliott et al., 2005 (30 invariant amino acids) and Kusch et al., 2016 (58 highly conserved amino

242 acids) were screened in all *CsMLO* sequences using our final protein alignment (**Supplementary**

243 **File S3**). Our manually curated *MLO* gene models in the five cannabis genomes were also screened

244 for conserved *cis*-acting elements in the promoter regions. A homemade Python script (v.3.7.3,

245 Van Rossum and Drake, 2009) was used to extract a 2 kb upstream region of each *CsMLO* gene,

246 and these promoter regions were used as search queries in the plantCARE database (Lescot et al.,

247 2002).

248

249 **2.3   Phylogenetic analysis**

250 Amino acid sequences for all manually curated CsMLO from CBDRx, Jamaican Lion female and

251 male, Purple Kush and Finola were aligned together with MLO sequences previously identified in

252 *Arabidopsis thaliana* (AtMLO, as indicated in Devoto et al., 2003), *Vitis vinifera* (grapevine:

253 VvMLO, as indicated in Feechan et al., 2008), *Prunus persica* (peach: PpMLO, as indicated in

254 Pessina et al., 2014), *Hordeum vulgare* (barley: HvMLO, as indicated in Kusch et al., 2016) and

255 *Zea mays* (maize: ZmMLO, as indicated in Kusch et al., 2016). *Chlamydomonas reinhardtii* MLO

256 (XP_001689918) was used as an outgroup. Alignment of protein sequences was performed using

257 MAFFT v7.407_1 (Katoh and Standley, 2013) with default parameters within NGPhylogeny.fr

258 (Lemoine et al., 2019) and used to construct phylogenetic trees. A first tree was constructed using

259 PhyML+SMS v1.8.1_1 (Lefort et al., 2017) with default parameters within NGPhylogeny.fr

260 (**Figure 3**), and a second tree was constructed using MrBayes v3.2.6_1 (Huelsenbeck and

261 Ronquist, 2001) with default parameters within NGPhylogeny.fr (**Supplementary Figure S1**).

262 The trees were interpreted and visualized using iTOL (Letunic and Bork, 2019). All MLO proteins

263 identified were classified into clades based on previous phylogenetic analyses (Kusch et al., 2016).

264

265 **2.4   Transcriptional activity of *CsMLOs* in response to powdery mildew infection**

266 **2.4.1   Sampling and RNA sequencing**

267   An RNA-seq time series analysis of the infection of cannabis by PM was performed to characterize

268   the transcriptional response of *CsMLO* genes. The experiment was conducted in a controlled

269   environment at Organigram Inc, a Health Canada approved licensed producer (Moncton, New-

270   Brunswick, Canada). Cannabis fan leaves from four-week-old vegetative plants ('*Pineapple*

271   *Express*', drug-type I) were manually inoculated with *G. ambrosiae* emend. (including *G.*

272   *spadiceus*) spores. Heavily infected leaves, loaded with fungal spores, were scraped against the

273   surface of the leaves from the four-week-old plants to induce infection. Leaf samples (punch holes)

274   were taken at five time points during the infection (n = 3 per time point): day zero (T0), six hours

275   post-inoculation (6H), one day post-inoculation (1D), three days post-inoculation (3D) and eight

276   days post-inoculation (8D). RNA samples were extracted from leaf tissues using the RNeasy®

277   Plant Mini Kit (QIAGEN, Hilden, Germany) following the standard manufacturer's protocol. Each

278   mRNA extraction was treated with QIAGEN's RNase-Free DNase Set (QIAGEN, Hilden,

279   Germany), involving a first round of the TURBO DNA-free™ DNA Removal Kit through the

280   extraction protocol and then two rounds of the DNA-free™ Kit (Life Technologies, Carlsbad, CA,

281   USA). Fifteen individual RNAseq libraries were generated for the five time points sampled (n = 3

282   per time point). cDNA libraries were sequenced on a total of six lanes, using the Illumina HiSeq

283   v4 technology (PE 125 bp) at the Centre d'expertise et de services Génome Québec (Montreal,

284   QC, Canada). In total, ~200 Gb of raw sequencing data were generated, which represents 1.728

285   billion of 2 x 125 bp paired-end sequences distributed across all 15 libraries (BioProject accession:

286   PRJNA738505, SRA accessions: SRR14839036-50).

287

288   **2.4.2   Short-read alignment on the reference genome and differential gene expression**

289   Raw sequencing reads were cleaned, trimmed, and aligned on the same reference genome as the

290   one initially used to find our final *MLO* gene models (CBDRx) to estimate changes in transcript-

291   specific levels of expression over the course of the infection. Specifically, mild trimming

292   thresholds were applied to clean and trim all raw reads, using Trimmomatic v.0.34 (Bolger et al.,

293   2014)   with   the   following   parameters: ILLUMINACLIP:$VECTORS:2:30:10,

294   SLIDINGWINDOW:20:2, LEADING:2, TRAILING:2, MINLEN:60. The cleaned reads from our

295   15 individual libraries were aligned on the CBDRx reference genome using STAR v.2.7.6 (Dobin

296   et al., 2013) in genome mode with default parameters and the official NCBI Cannabis annotation

297   release 100 associated with the genome. Genome-wide raw read counts were obtained for each

298   library using htseq-count v.0.11.1 (Anders et al., 2015) with the 'intersection-non-empty' mode.

299

300   Downstream analyses of differential gene expression patterns were conducted using the R

301   packages 'limma-voom' (Law et al., 2014) and edgeR (Robinson et al., 2010) in RStudio

302   v.1.3.1073 (RStudio Team, 2020). Reference sequences with insufficient sequencing depth were

303   filtered out by keeping only the ones with more than five Counts Per Million (CPM) in at least

304   three samples. This mild CPM threshold allowed the filtering of very low coverage genes without

305   losing too much information in the dataset. This filtered dataset was normalized using the Trimmed

306   Mean of M-values (TMM) method implemented in edgeR. A transformation of the data was then

307   performed using the 'limma-voom' function, which estimates the mean-variance relationship for

308   each transcript, allowing for better and more robust comparisons of gene expression patterns across

309   RNA-seq libraries (Law et al., 2014). Each transcript was finally fitted to an independent linear

310   model with log2(CPM) values as the response variable and the time point (zero, six hours, three

311   days, and eight days post-infection) as the explanatory variable. All linear models were treated

312   with limma's empirical Bayes analysis pipeline (Law et al., 2014). Differentially expressed genes

313   were chosen based on a False Discovery Rate (FDR, Benjamini–Hochberg procedure) < 0.01.

314   Genomic regions corresponding to the curated CBDRx *MLO* gene models were extracted from the

315   output of edgeR/limma-voom for each comparison made (i.e. each time point compared to T0) and

316   looked for significant gene expression differences among *MLO* genes (FDR < 0.01). These

317   expression differences in *MLOs* over the course of the infection were visualized in RStudio

318   v.1.3.1073 (RStudio Team, 2020) on a scatter plot using CPM values (**Figure 4, Supplementary**

319   **Figure S2**).

320

321   **2.5   Cloning of clade V *CsMLOs* for transient expression in *Nicotiana benthamiana* and**

322   **confocal microscopy**

323   The two selected *MLO* gene sequences (*CsMLO1* and *CsMLO4*) were synthesized commercially

324   into the Gateway-compatible vector pDONR™/Zeo (Invitrogen, Carlsbad, CA, USA) (Hartley et

325   al., 2000). The entry vectors were inserted into *Escherichia coli* OneShot® TOP10 cells

326   (Invitrogen, Carlsbad, CA, USA) by chemical transformation according to the manufacturer's

327   instructions. Positive colonies were selected, and plasmid DNA was extracted with the EZ-10 Spin

328   Column Plasmid DNA Miniprep Kit (Bio Basic Inc., Markham, ON, Canada). The extracted entry

329   vectors were confirmed by PCR with the primers M13-F (5'-GTAAAACGACGGCCAGT-3') and

330   M13-R (5'-CAGGAAACAGCTATGAC-3') as well as M13-F and MLO1_1-R (5'-

331   ATGTGCCATTATAAATCCATGCCT-3', this study).

332

333   The Gateway-compatible destination vector chosen was pB7FWG2.0, which is under the

334   regulation of the 35S Cauliflower Mosaic Virus (CaMV) promoter and harbors the plant selectable

335   marker gene *bar* (bialaphos acetyltransferase), which confer resistance against glufosinate

336   ammonium. It also possesses a streptomycin and spectinomycin resistance gene for plasmid

337   selection, and an EGFP-fusion in C-terminal, for visualization by confocal microscopy (Karimi et

338   al., 2002). According to the manufacturer's instructions, the destination vector was inserted into

339   *Escherichia coli* One Shot® ccdB Survival™ cells (Life Technologies) by chemical

340   transformation. Positive colonies were selected, and plasmid DNA was extracted with the EZ-10

341   Spin Column Plasmid DNA Minipreps Kit (Bio Basic Inc., Markham, ON, Canada). The extracted

342   destination vectors were confirmed by PCR with the primers T-35S-F (5'-

343   AGGGTTTCTTATATGCTCAACACATG-3', Debode et al., 2013) and EGFP-C (5'-

344   CATGGTCCTGCTGGAGTTCGTG-3').

345

346   The two synthesized MLO gene sequences were then inserted into the vector pB7FWG2.0

347   through an LR clonase reaction following the manufacturer's instructions (Invitrogen, Carlsbad,

348   CA, USA). Plasmids were then transferred to *E. coli* OneShot® TOP10 cells (Invitrogen, Carlsbad,

349   CA, USA) by chemical transformation according to the manufacturer's instructions. Positive

350   colonies were selected, and plasmid DNA was extracted with the EZ-10 Spin Column Plasmid

351   DNA Minipreps Kit (Bio Basic Inc., Markham, ON, Canada). The extracted expression vectors

352   were amplified by PCR with the primers 35S Promoter (5'-CTATCCTTCGCAAGACCCTTC-3')

353   and MLO1_1-R as well as MLO2_3-F (5'-TCTTTCAGAATGCATTTCAACTTGC-3', this study)

354   and EGFP-N (5'-CGTCGCCGTCCAGCTCGACCAG-3'). Sequencing of the PCR products above

355   was performed to confirm the junction between the plasmid and the inserted gene and the junction

356   between the inserted gene and the GFP, respectively.

357

358   The recombinant vectors were then transferred to *Agrobacterium tumefaciens*

359   ElectroMAX™ LBA4404 cells by electroporation (Invitrogen, Carlsbad, CA, USA). The

360   expression vector in *A. tumefaciens* was confirmed by PCR using the primers 35S Promoter and

361   MLO1_1R and MLO2_3-F + EGFP-N. Cultures of transformed *A. tumefaciens* (*CsMLO1* and

362   *CsMLO4*) were incubated with agitation at 28°C in LB broth containing spectinomycin (100

363   mg/mL) for 24 hours. The cultures were then centrifuged at 5000 x *g* for 5 minutes, the supernatant

364   was discarded, and the pellet was resuspended in MgCl2 (10 mM). The cells were brought to an

365   OD600 of 0.5 and incubated at room temperature for 2 hours with acetosyringone (200 µM).

366   *Nicotiana benthamiana* plants, about two weeks old, were watered a few hours before infiltration,

367   and the bacterial suspensions were administered using sterile 1 mL syringes (without needles) on

368   the abaxial surface of the leaves. The plants were then returned to growth chambers, and the

369   observation of epidermal cells was performed three days after infiltration using a confocal laser

370   scanning microscope. Leaves were observed under a Leica TCS SP8 confocal laser scanning

371   microscopy (Leica Microsystems). Images were observed through an HC PL APO CS2 40X/1.40

372   oil immersion objective at excitation/emission wavelengths of 488/503-521 nm.

373

374   **2.5   *In silico* screening of clade V *CsMLO* sequence variants in 32 cannabis cultivars**

375   Thirty-one distinct drug-type *Cannabis sativa* cultivars from Organigram Inc. (Moncton, New-

376   Brunswick, Canada) and one industrial hemp variety ('Anka', UniSeeds, obtained from Céréla,

377   Saint-Hughes, Québec, Canada) were screened to identify potential polymorphisms in clade V

378   *CsMLOs* that could be associated with increased susceptibility or resistance to PM. Raw

379   sequencing files (Illumina paired-end 125 bp) from these 32 cultivars (BioProject accession:

380   PRJNA738519, SRA accessions: SRR14857079-110) were aligned on the reference CBDRx

381   genome using 'speedseq' v.0.1.2 (Chiang et al., 2015). 'bcftools view' v.1.10.2 (Li, 2011) was

382   used on the raw BAM alignment files to extract the genomic regions corresponding to the two

383   clade V *CsMLO* genes, based on the positions of our manually curated CBDRx *MLO* gene models.

384   Genotypes in these two *CsMLOs* were called for Single Nucleotide Polymorphisms (SNPs) and

385   small insertions and deletions (INDELs) across the 32 cannabis cultivars using 'bcftools mpileup'

386   v.1.10.2 (Li, 2011). Variants with a mapping quality < 20 and/or with a read depth > 500X were

387   filtered out, and allelic frequencies for each variant were extracted using an in-house Python script

388   (v.3.7.3, Python Software Foundation, 2020). Next, SNPGenie v.1.0 (Nelson et al., 2015) was used

389   to estimate the ratio of non-synonymous to synonymous polymorphisms in the two clade V

390   *CsMLO* genes across all 32 cultivars. Based on the output from SNPGenie, genomic positions of

391   all non-synonymous polymorphisms found in the two genes were extracted using an in-house

392   Python script (v.3.7.3, Python Software Foundation, 2020). Visual representations of these

393   polymorphisms at the gDNA and amino acid levels were prepared using Microsoft © PowerPoint

394   v.16.49 (Microsoft Corporation, Redmond, Washington, U.S.).

395

396   **3**   **Results**

397

398   **3.1**   ***CsMLO* gene identification and genomic localization**

399   Through careful manual curation, we were able to identify a total of 14, 17, 19, 18 and 18 distinct

400   *CsMLO* genes in the genomes of CBDRx, Jamaican Lion female, Jamaican Lion male, Purple

401   Kush, and Finola, respectively. Our final manually curated *CsMLO* genes were numbered 1-15

402   based on chromosomal positions in CBDRx, from chromosome 1 through chromosome X (**Table**

403   **1**, **Supplementary Tables S1-S5, Supplementary Files S1-S3**). *CsMLO* gene numbers and IDs

404   in all four other genomes (Finola, Jamaican Lion female, Jamaican Lion male, Purple Kush) were

405   based on homology relationships, supported by our phylogenetic and orthology analyses (see

406   Materials and Methods for details). We also physically located each set of *CsMLO* genes on the

407   chromosomes of CBDRx, Finola and Purple Kush (**Figure 1**). Our results show that eight of the

408   10 chromosomes harbor evenly-spaced *CsMLO* genes, chromosomes 5 and 9 being the only ones

409   not carrying MLO genes. We were able to anchor all 14 CBDRx *CsMLO* genes on their respective

410   chromosomes, while two (*CsMLO12-FN*, *CsMLO15-FN*) and six (*CsMLO1-PK_B*, *CsMLO2-PK*,

411   *CsMLO4-PK*, *CsMLO6-PK*, *CsMLO7-PK*, *CsMLO12-PK*) *MLO* genes were located on

412   unanchored contigs in Finola and Purple Kush, respectively (**Figure S1, Table 1, Supplementary**

413   **Tables S1, S4 and S5**). The CBDRx genome showed an absence of MLO15, while this gene was

414   present in all other cannabis genomes. We identified homology and paralogy relationships among

415   all *CsMLO* sequences, which revealed potential duplication patterns in certain genomes. Overall,

416   we detected paralogs for *CsMLO1*, *CsMLO5*, *CsMLO9*, *CsMLO10*, *CsMLO12* and *CsMLO13* in

417   the genomes of Finola, Purple Kush, Jamaican Lion male and Jamaican Lion female (*CsMLO*

418   paralogs were designated with A/B suffixes, see **Table 1, Supplementary Tables S1-S5**). CBDRx

419   remained the only genome in which we did not identify *CsMLO* paralogs. Most of the *CsMLO*

420 genes had syntenic positions across all three chromosome-level genome assemblies, with the
421 exception of *CsMLO9*, *CsMLO10* and *CsMLO14* (**Figure 1**). *CsMLO14* was the only *CsMLO* gene
422 identified across all five genomes that had three different locations in the three genomes, i.e.
423 chromosome X in CBDRx, chromosome 1 in Finola and chromosome 4 in Purple Kush.
424 Partial/incomplete genes were also noted in the genomes of Finola (*CsMLO3-FN-B* and
425 *CsMLO10-FN-A*) and Purple Kush (*CsMLO1-PK-B*, *CsMLO2-PK* and *CsMLO5-PK-A*).

426

427 **3.2**   ***CsMLO* gene structure and protein characterization**

428 Manually curated *CsMLO* genes identified in CBDRx ranged in size between 3,567 bp
429 (*CsMLO14-CBDRx*) and 49,009 bp (*CsMLO5-CBDRx*), with an average size of 11,240 bp and a
430 median size of 5,431 bp (**Table 1, Supplementary Tables S1-S5**). The structural organization of
431 these *CsMLO* genes is depicted in **Figure 2**. The number of exons varied between 12 and 15, with
432 some of the exons showing signs of fusion in certain genes (**Figure 2**). The number of amino acid
433 residues in these *CsMLO* gene sequences varied between 509 and 632. Intron size varied
434 considerably, with 11 introns belonging to four different *CsMLO* genes (*CsMLO4*, *CsMLO5*,
435 *CsMLO6*, *CsMLO10*) exhibiting a length greater than 1,000 bp. *CsMLO5* had the longest introns:
436 intron 5 (15.6 kb), intron 9 (7.9 kb), intron 12 (15.9 kb), and intron 14 (6.3 kb) (**Figure 2**). The
437 *CsMLO* genes that we characterized in the four other genomes (Finola, Jamaican Lion female and
438 male, Purple Kush) were consistently similar to the ones identified in CBDRx in terms of length,
439 intron and exon structural organization and genomic localization (**Supplementary Files S1-S3**).
440 The longest *CsMLO* gene characterized among all our manually curated gene models belonged to
441 Jamaican Lion (female), with a total length of 49,673 nucleotides (*CsMLO5-JL*).

442

443 Proteins encoded by all identified *CsMLO* genes comprised seven transmembrane domains
444 (TMs) of similar lengths (**Table 1, Supplementary Tables S1-S5**). The only exceptions were
445 found in the genomes of Finola and Purple Kush in which we identified a total of four partial
446 *CsMLO* genes, which harbored less than seven TMs: CsMLO3-FN-B (5 TMs), CsMLO10-FN-A
447 (3 TMs), CsMLO1-PK-B (4 TMs), CsMLO5-PK-A (6 TMs). Similarly, all identified *CsMLO*
448 proteins were predicted to be localized in the cell membrane by several online prediction servers,
449 such as Plant-mPLoc (Chou and Shen, 2010), DeepLoc 1.0 (Almagro Armenteros et al., 2017),
450 and YLoc-HighRes (Briesemeister et al., 2010) (**Table 1, Supplementary Tables S1-S5**). Three

451 partial/incomplete sequences were also predicted to localize elsewhere, such as the chloroplasts

452 and nucleus by Plant-mPLoc for CsMLO1-PK-B, and the endoplasmic reticulum by DeepLoc 1.0

453 for CsMLO5-PK-A and CsMLO3-FN-B. We assumed that these three predictions were unreliable

454 as they were made using incomplete sequences and not present unanimously throughout all

455 prediction servers. No signal peptide (SignalP 5.0, Almagro Armenteros et al., 2019a),

456 mitochondrial transit peptide, chloroplast transit peptide, or thylakoid luminal transit peptide

457 (TargetP 2.0, Almagro Armenteros et al., 2019b) were predicted in any of the CsMLO protein

458 sequences (results not shown). The invariable 30 amino acid residues previously described in

459 Elliott et al., 2005 were identified in all CsMLOs (**Table 1, Supplementary Tables S1-S5**).

460 Across all identified CsMLOs (excluding the five partial sequences), the amount of conserved

461 amino acids varied between 25 and 30, and 74.1% (60/81) of CsMLO sequences possessed all 30

462 amino acids. In Kusch et al., 2016, a larger dataset of MLO proteins was analyzed and thus

463 identified 58 highly conserved amino acids, rather than invariant, showing that substitutions are

464 possible. These 58 amino acid residues were also screened in all CsMLOs (**Table 1,**

465 **Supplementary Tables S1-S5**). Across all identified CsMLOs (excluding the five partial

466 sequences), the amount of conserved amino acids varied between 52 and 58, and only 23.5%

467 (19/81) of CsMLO sequences possessed all 58 amino acids. However, 74.1% (60/81) possessed

468 57 or more of the conserved amino acids.

469

470 **3.3   Phylogenetic analysis of CsMLOs**

471 We performed a phylogenetic analysis on the curated cannabis MLO proteins identified among the

472 five genomes (CsMLOs). The dataset was completed with the MLO protein family from

473 *Arabidopsis thaliana* (AtMLO, Devoto et al., 2003), *Prunus persica* (PpMLO, Pessina et al.,

474 2014), *Vitis vinifera* (VvMLO, Feechan et al., 2008), *Hordeum vulgare* and *Zea mays* (HvMlo and

475 ZmMlo, Kusch et al., 2016), using the *Chlamydomonas reinhardtii* MLO as the outgroup.

476 Phylogenetic tree construction was performed using the PhyML+SMS algorithm, which confirmed

477 the seven known clades of MLO proteins (**Figure 3**), with bootstrap values equal or greater than

478 97% (except for clade I, supported with a bootstrap value of 77%). Clade numbers from I to VII

479 were assigned according to the previous study of Kusch et al. (2016). Previous studies have

480 reported the presence of an eighth clade (e.g., Pessina et al., 2014), clustering with clade VII in

481 other papers. Here, we followed the seven clades nomenclature, but this potential eighth clade

482   would correspond to one of the two clade VII subclades. Indeed, potential subclades were also

483   identified in our study, indicated as separate lines in the inner circle of **Figure 3**, which were also

484   supported with high values of bootstrap (equal or greater than 83%, with the exception of one clade

485   V subclade, supported with a value of 67%). The same clades and subclades were also found to be

486   supported by high posterior values equal or greater than 77% (indicated in red), following

487   phylogenetic tree construction using MrBayes (also see **Supplementary Figure S1**). Apart from

488   one clade II subclade that appeared monocot-specific, all subclades depicted here included

489   CsMLO, as well as PpMLO and VvMLO. In each subclade, all CsMLOs clustered together with

490   bootstrap values of 100. Two subclades were found in the phylogenetic clade V, containing all the

491   dicot MLO proteins experimentally shown to be required for PM susceptibility (Acevedo-Garcia

492   et al., 2014). In one of those clade V subclades, three of the cannabis genomes were found to harbor

493   two near-identical genes. All cannabis genomes except CBDRx were found to include one *MLO*

494   gene grouping with clade IV, which contains all monocot MLO proteins acting as PM

495   susceptibility factors (**Figure 3**).

496

497   **3.4   Transcriptional reprogramming of cannabis *MLOs* in response to *Golovinomyces***

498   ***ambrosiae* infection**

499   We conducted an RNA-seq analysis to look at the transcript abundance of *CsMLO* genes in leaves

500   of the susceptible cannabis cultivar "Pineapple Express" during infection by *G. ambrosiae* emend.

501   (including *G. spadiceus*). Five time points were investigated, corresponding to key stages of the

502   infection: 0 (control), 6 hours post-inoculation (hpi) (conidia germination and appressoria

503   formation), 24 hpi (haustoria formation), 3 day post-inoculation (dpi) (secondary hyphae

504   formation), and 8 dpi (secondary haustoria, secondary appressoria and conidiophore/conidia

505   formation). No significant expression was detected at any time point for *CsMLO6*, *CsMLO8*,

506   *CsMLO11*, and *CsMLO15* (in this particular case, reads were aligned to the genome of Jamaican

507   Lion male, as the CBDRx genome is devoid of *CsMLO15*). While *CsMLO3* and *CsMLO11* were

508   expressed, no up- or down-regulation was observed under our conditions. Nine genes, namely

509   *CsMLO1*, *CsMLO2*, *CsMLO4*, *CsMLO5*, *CsMLO7* and *CsMLO9* were found to be significantly

510   differentially expressed (FDR < 0.05) after inoculation with the pathogen (**Figure 4,**

511   **Supplementary Figure S2**). In the case of clade V genes, *CsMLO1* showed a peak of 2.23-fold

512   up-regulation at 6 hpi (FDR = 2.83×10^{-4}), and remained somewhat constant for the remaining of

513   the infection, while the expression of *CsMLO4* increased steadily between T0 and 3 dpi, reaching

514   a peak of 2.02-fold up-regulation (when compared to T0) and remained constant at 8 dpi (FDR =

515   7.87×10^{-4}).

516

517   An analysis of the 2 kb upstream region of all *CsMLO* genes identified through the five

518   cannabis genomes revealed the presence of key regulatory motifs with functions related to

519   environmental/hormonal response (e.g., ABRE, AuxRR-core, ERE, GARE, P-box, TATC-box,

520   TGA-element, GT1-motif, G-box, light response elements), stress and defense response (TC-rich

521   repeats, MBS, ARE, GC-motif, LTR element), developmental regulation (circadian, HD-Zip,

522   CCAAT-box, MSA-like), seed-specific metabolism (O2-site, RY-element) and wound response

523   (WUN-motif). In total, 82 (95%) *CsMLO* genes had at least one motif related to

524   environmental/hormonal response, while 80 (93%) *CsMLO* genes had a MYB-related sequence, a

525   motif typically involved in development, metabolism and responses to biotic and abiotic stresses

526   (Dubos et al., 2010) (**Supplementary Table S6**). We found the presence of at least one *cis*-acting

527   element involved in gene overexpression by biotic and abiotic factors (ABRE, CGTCA, TGACG,

528   TCA) in 100% of the 13 *CsMLO* genes from clade V identified in all five genomes. The analysis

529   of protein domains, their location in the protein and the overall topology of each gene ultimately

530   revealed a consistent pattern among all *CsMLO* genes.

531

532   **3.5   Subcellular localization of clade V *CsMLOs***

533   The subcellular localization of clade V MLOs, such as CsMLO1 and CsMLO4, was first analyzed

534   using online tools. As mentioned previously, these analyses predicted that CsMLO1 and CsMLO4

535   possessed seven transmembrane domains and were localized in the plasma membrane (**Table 1,**

536   **Supplementary Tables S1-S5**). To determine the subcellular localization of CsMLO1 and

537   CsMLO4 *in planta*, we constructed two vectors under the control of the CaMV 35S promoter

538   where the coding sequences of CsMLO1 and CsMLO4 were fused to enhanced green fluorescent

539   protein (EGFP) in C-terminal (*35S::CsMLO1-EGFP* and *35S::CsMLO4-EGFP*). The

540   agroinfiltration-based transient gene expression system was used to transform *Nicotiana*

541   *benthamiana* leaves with each construct. The epidermal cells were observed three days after

542   transformation for GFP signal using confocal laser scanning microscopy. The CsMLO1-EGFP

543   fusion protein was observed in the cell periphery as well as throughout the cell forming networks

544 in a punctuate pattern, while the CsMLO4-EGFP fusion protein was observed solely in the cell

545 periphery in a defined way (**Figure 5**).

546

547 **3.6 Polymorphism analysis of clade V *CsMLOs* among 32 cultivars**

548 Comparison of 32 distinct cannabis cultivars to the CBDRx reference genome to detect

549 polymorphisms in *CsMLO* clade V genes revealed a total of 337 (*CsMLO1*, 101 indels and 236

550 SNPs) and 852 (*CsMLO4*, 154 indels and 698 SNPs) polymorphisms, mainly located in the last

551 cytosolic loop of the protein, near the C-terminus. Among these polymorphic loci, we identified

552 14 and 12 non-synonymous SNPs for *CsMLO1* and *CsMLO4* respectively (**Supplementary Table**

553 **S7**). The only SNP that was not located near the C- or N-terminus was found in *CsMLO1*, in the

554 second cytoplasmic loop. All other SNPs in *CsMLO1* were distributed in the first cytoplasmic loop

555 (three non-synonymous SNPs), the second extracellular loop (four non-synonymous SNPs) and

556 the last cytoplasmic loop, near the C-terminus (six non-synonymous SNPs, **Figure 6**). The

557 scenario for *CsMLO4* is slightly different, with a single non-synonymous SNP identified in the

558 second extracellular loop and the remaining 11 SNPs all located in the last cytoplasmic loop, near

559 the C-terminus (**Figure 6**). About half of these non-synonymous SNPs in both genes induced a

560 change in amino acid charges or polarity, with seven (50%) and five (42%) SNPs having a change

561 in electric charges in *CsMLO1* and *CsMLO4*, respectively. None of the SNPs identified in either

562 of the two *CsMLO* clade V genes had an impact on conserved amino acids in the proteins (Elliott

563 et al., 2005; Kusch et al., 2016). Overall, allele frequencies associated with these SNPs in *CsMLO1*

564 and *CsMLO4* showed an even distribution of reference and alternate alleles throughout the 32

565 cultivars. There was one exception with 'Ultra Sour', which was identified as the only homozygous

566 cultivar for the alternate allele in three non-synonymous SNPs found in exons 1 (G40E) and 3

567 (P96Q, P111T).

568

569 **4 Discussion**

570

571 Considering the role of specific *MLO* genes in flowering plants' susceptibility to powdery mildew,

572 one of the most prevalent pathogens in indoor cannabis productions (Punja et al., 2019), our

573 primary goal was to structurally and functionally characterize this gene family at a manual-curation

574 level in multiple cannabis genomes. In order to develop mitigation strategies aimed at reducing

575   the deleterious impacts of the pathogen on cannabis productions, and in an attempt to better

576   understand other functional roles of *CsMLOs*, the first step consisted in identifying the exact

577   number and structure of these genes in different genetic backgrounds. Our results first showed that

578   *CsMLO* numbers are variable across different cannabis types. Second, they showed that two

579   distinct clade V genes were present in all genomes (with paralogs in certain cultivars) and that

580   these clade V genes possessed *cis*-acting elements typically overexpressed by biotic and abiotic

581   factors. These specific elements in the promoter regions of clade V *CsMLOs* allow them to be

582   responsive to experimental PM infection, which we validated in the context of an infection time

583   point experiment.

584

585   **4.1   *CsMLOs* and the importance of manual curation**

586   We used 15 *Arabidopsis thaliana* MLO protein sequences to mine the genomes of five cannabis

587   cultivars of four different types (THC-dominant, balanced THC:CBD, CBD-dominant, food-oil

588   hemp), yielding a sum of 86 *CsMLOs* across all genomes (**Table 1, Supplementary Tables S1-**

589   **S5**). According to our data, these genes are organized into 15 *CsMLO* homologs in total (**Figures**

590   **1 and 3, Supplementary Tables S1-S5**). We were not able to retrieve *CsMLO15* in CBDRx,

591   making this genome devoid of clade IV *MLO*. Whether this is an artifactual gene loss resulting

592   from the genome assembly and cleaning process or a biological reality in this specific cultivar

593   remains to be verified with additional sequencing data. If this is a biological reality, it could suggest

594   that these genes potentially have redundant and similar functions or are imbricated into functional

595   networks buffered by redundancy (Tully et al., 2014; AbuQamar et al., 2017). The loss of a gene

596   in the genome of CBDRx in this context could potentially be phenotypically less detrimental. Gene

597   length also varied with a 14-fold difference between the shortest and longest sequence, with two

598   genes (*CsMLO5* and *CsMLO10*) exhibiting multiple unusually long introns (> 10,000 nucleotides,

599   **Figure 2**). Intron length distribution across the 15 *CsMLOs* showed considerable variability, which

600   explained the significant variations observed in gene length across all *CsMLOs*. Plant introns are

601   typically relatively short and rarely extend beyond 1 kb, making these large *CsMLO* introns up to

602   10 times longer than typical plant exons (Wu et al., 2013). However, the numbers and positions of

603   exons and introns for the same homologs across the five genomes were highly conserved. We

604   found that cannabis genes tend to have on average five introns (median number of introns across

605   the genome = 4), indicating that *CsMLOs* have three times the number of introns found in a typical

606 gene from the cannabis genome. This could indicate that those introns are likely to play an
607 important functional role and, thus, may be a significant aspect of gene regulation (Seoighe and
608 Korir, 2011). On the other hand, selection against intron size could be counterbalanced in *CsMLOs*
609 by a selective preference for larger introns which correlates with more regulatory elements and a
610 more complex transcriptional control (e.g., in *Vitis vinifera*, Jiang and Goertzen, 2011). Even
611 though the complete *CsMLO* gene catalog could be retrieved in each genome through gene
612 prediction algorithms combined with targeted BLAST searches, the precise characterization of
613 each gene structure (i.e. start/stop codons, coding sequence, exon-intron boundaries) could not be
614 achieved without multiple efforts of manual curation.

615

616 Automated gene structure prediction algorithms are often considered sufficiently reliable
617 to recover the complete genome-wide repertoire of genes of a given sample. However, as repeat
618 content, size, and structural complexity of those genes increase (i.e. numerous small exons
619 delimited by long introns, as observed in *CsMLOs*), errors are increasingly likely to occur and thus
620 impair the accuracy of automated annotations (Guigó et al., 2000; Pilkington et al., 2018). Because
621 13% of the *CsMLOs* identified here had introns larger than 10 kb (*CsMLO5* and *CsMLO10*), and
622 two additional genes (*CsMLO4* and *CsMLO6*) had introns larger than 1,000 bp, most of these
623 *CsMLOs* were mispredicted by automated gene prediction algorithms (**Supplementary Tables**
624 **S1, S2 and S3**). These algorithms typically have 10 kb as the default maximum intron length (e.g.,
625 in MAKER2, Holt and Yandell, 2011). In comparison, *Arabidopsis thaliana* genome annotation
626 version 10 indicates that there are 127,854 introns in the nuclear genes, and of these, 99.23% are
627 less than 1,000 bp, while only 16 introns are larger than 5 kb (NCBI accession: GCF_000001735.4,
628 TAIR10.1). Long multi-exon genes having long introns end up fragmented into several shorter
629 "genes" by these programs, thus inflating the actual number of genes within the family. The
630 severity of such errors is influenced by various factors such as the quality of the assembly (Yandell
631 and Ence, 2012; Treangen and Salzberg, 2012) and the availability and quality of extrinsic
632 evidence (e.g., RNA-seq, orthologous sequences). While assembly quality is influenced by
633 genome size and repeat content (Tørresen et al., 2019; Whibley et al., 2021), the disparity in the
634 number of mispredicted genes observed in this study is also likely to be related to differences in
635 sequencing technology (Oxford Nanopore vs. Pacific Biosciences), sequencing depth, algorithms
636 used for *de novo* assembly, and simple base calling accuracy. Overall, these results revealed that

637   cannabis possesses an extensive repertoire of *MLOs* characterized by significant gene size

638   variations across all family members. These results also demonstrate the importance of manual

639   curation when working with automatically generated gene models. Identifying the complete and

640   detailed set of *CsMLOs* for each genome allowed the possibility to assess synteny and evolutionary

641   patterns among *Cannabis sativa* and other plant species.

642

643   **4.2   Evolutionary dynamics of *CsMLOs*, gene duplications and potential implications**

644   The vast majority of studies investigating phylogenetic relationships within the *MLO* gene family

645   usually classifies its members in seven clades. A few times, an extra clade has been proposed, or

646   various subclades have been identified, but no consensus has been reached yet. In our study, we

647   classified genes according to the seven clades, and identified some potential subclades presented

648   in **Figure 3**, most of which appear concordant with results from other studies. In two previous

649   studies (Polanco et al., 2018; Rispail and Rubiales, 2016), clade I is divided in two subclades (Ia

650   and Ib). Subclade Ia would correspond in our tree to the two subclades respectively including

651   *CsMLO12* and *CsMLO9*, while subclade Ib would correspond to the subclade that includes

652   *CsMLO2*. However, Iovieno et al. (2016) divided clade I in three subclades (Ia, Ib and Ic).

653   Subclade Ia from Iovieno et al. (2016) appears to correspond to subclade Ib from Polanco et al.

654   (2018), and in our tree is represented by the subclade including *CsMLO2*. Subclades Ib and Ic from

655   Iovieno et al. (2016) would correspond to subclade Ia from Polanco et al. (2018) and would be

656   represented by the subclades including *CsMLO12* and *CsMLO9*, respectively. In Iovieno et al.

657   (2016), clade II is divided into 15 subclades (named from IIa to IIq excluding IIj). Subclades IIa

658   and IIb would be represented in our tree by the subclade including *CsMLO6*; subclades IIc, IId and

659   IIe would be represented by the monocot-specific subclade; subclades IIf, IIg and IIh would be

660   represented by the subclade including *CsMLO13*; and subclades IIi to IIq would be represented by

661   the subclade including *CsMLO14*. Still in Iovieno et al. (2016), clade III is divided into 3 subclades

662   (IIIa, IIIb and IIIc), subclades IIIb and IIIc corresponding in our tree to the subclades including

663   *CsMLO8*, and *CsMLO7*, respectively. Subclade IIIa would correspond to the two immediate

664   outlying monocot sequences (*ZmMlo2* and *ZmMlo3*), while our two next outlying monocot

665   sequences (*HvMlo2* and *ZmMlo4*) would correspond to clade VIII (in Iovieno et al., 2016, as there

666   is no consensus on clade VIII). Clade IV has also been divided into two subclades (IVa and IVb)

667   which are grouped together in our tree, subclade IVa from their study simply corresponding to

668 monocot sequences and subclade IVb corresponding to dicot sequences. Clade V has been divided
669 into 3 subclades (Va, Vb and Vc), for which subclade Va corresponds in our tree to the subclade
670 including *CsMLO1*, while subclade Vb would correspond to the two outlying sequences from the
671 other subclade from our tree (*PpMLO3* and *VvMLO3*), and subclade Vc would correspond to the
672 remaining of this subclade that includes *CsMLO4*. Clade VI is not divided into subclades in this
673 paper, while we clearly identified two subclades in our tree, one including *CsMLO5* and one
674 including *CsMLO3*. Clade VII has been divided into two subclades in Iovieno et al. (2016), and
675 the same division can be found in our tree, with Iovieno's subclade VIIa corresponding to the
676 subclade that includes *CsMLO11*, while subclade VIIb would correspond to the subclade including
677 *CsMLO10*. In Zheng et al. (2016), a different clade VIII had been defined, which would correspond
678 to Iovieno et al. (2016) subclade VIIa, while Zheng et al. (2016) clade VII would correspond to
679 Iovieno et al. (2016) subclade VIIb. In most studies, this eighth clade has been merged within clade
680 VII, which is also the case here. As described above, a distinct clade VIII was also defined as a
681 monocot-specific clade in Iovieno et al. (2016), making the use of an 8-clades system confusing.
682 In our opinion, the clade VIII described by Iovieno et al. (2016) could be considered as a subclade
683 of clade III, represented in our tree by the most "diverged" sequences in this clade, *HvMlo2* and
684 *ZmMlo4*.

685

686   Manual curation of *CsMLOs* across the five studied genomes revealed with exactitude their
687 respective genomic localization, showing an overall conserved syntenic pattern, except for two
688 genomes, Finola and Purple Kush, which exhibited distinct chromosome localizations for specific
689 *CsMLO* orthologs, as compared to the rest of the genomes. The number of *CsMLOs* per genome,
690 ranging from 14 to 18, was comparable to other plant genomes, such as *Arabidopsis thaliana* (15,
691 Chen et al., 2006), *Vitis vinifera* (14, Feechan et al., 2008), *Cucumis sativus* (13, Zhou et al., 2013),
692 *Solanum lycopersicum* (15, Chen et al., 2014), *Hordeum vulgare* (11, Kusch et al., 2016),
693 *Medicago truncatula* (14, Rispail and Rubiales, 2016), *Cicer arietinum* (13, Rispail and Rubiales,
694 2016), *Lupinus angustifolius* (15, Rispail and Rubiales, 2016), *Arachis* spp. (14, Rispail and
695 Rubiales, 2016), *Cajanus cajan* (20, Rispail and Rubiales, 2016), *Phaseolus vulgaris* (19, Rispail
696 and Rubiales, 2016) and *Vigna radiata* (18, Rispail and Rubiales, 2016). The genomes of Finola
697 and Purple Kush, however, exhibited certain anomalies. Firstly, we found a greater proportion of
698 genes located on unanchored contigs in Finola (11%) and Purple Kush (33%) as compared to

699   CBDRx (0%). Secondly, two paralog pairs, one in Finola (*CsMLO10-FN-A* and *CsMLO10-FN-B*)

700   and one in Purple Kush (*CsMLO9-PK-A* and *CsMLO9-PK-B*), had one of the two paralogs located

701   on a different chromosome (**Figure 1**).

702

703   The genome-wide distribution of *CsMLOs* described here did not suggest the involvement

704   of tandem duplications as a predominant mechanism of emergence for *MLOs* in cannabis, as

705   suggested in other taxa (e.g., Liu and Zhu, 2008; Pessina et al., 2014; Rispail and Rubiales, 2016).

706   Indeed, recent bioinformatic analyses suggested that segmental and tandem duplications were a

707   widespread mechanism for the expansion of the *MLO* gene family in diverse plant species,

708   spanning from algae to dicots (Shi et al., 2020). For example, clear evidence of tandem duplication

709   events have been detected in *P. vulgaris* and *V. radiata*, and in *M. domestica*, respectively (Rispail

710   and Rubiales, 2016; Pessina et al., 2014). We did not find evidence that tandem duplications were

711   widespread in *CsMLOs*, as most of the genes were evenly spread out in the genomes, with multiple

712   other unrelated genes in between these *CsMLOs*. Some of the *CsMLOs* were located physically

713   close to one another and they were genetically related, but not similar enough ($< 85\%$) to be

714   considered as tandem duplicates. In this case, segmental duplication appears to be a more likely

715   mechanism of emergence for *CsMLOs*, although we did not specifically search for segmental

716   duplications in the present study. In total, four (4.7%) *CsMLOs* across the five genomes were

717   located in tandem duplications (**Figure 1**). These four putative tandem gene duplications were

718   located in Finola only, on chromosome X (*CsMLO13-FN-A* and *CsMLO13-FN-B*) and on

719   chromosome 2 (*CsMLO3-FN-A* and *CsMLO3-FN-B*). Other *CsMLOs* that could represent

720   potential tandem gene duplications in Jamaican Lion male (*CsMLO1-JLm*, *CsMLO12-JLm,*

721   *CsMLO13-JLm*) and female (*CsMLO1-JL*, *CsMLO10-JL*) all had the two paralogs located on

722   different contigs that were, on average, longer than 2 Mb each. These different contigs containing

723   two *CsMLO* paralogs typically harbored large sequences (> 10 kb) of high homology ($> 95\%$

724   similarity), which could suggest that they are either the result of segmental duplications, or that

725   they represent two copies of highly polymorphic loci (Fan et al., 2008; Lallemand et al., 2020).

726   We did not find evidence of tandem duplications in the genome of Purple Kush, indicating that

727   this genome is potentially more fragmented than the others. The duplication of *CsMLO13* (clade

728   II) on chromosome X is of potential interest as it was duplicated in the male genomes only (Finola

729   and Jamaican Lion male, **Supplementary Tables S3 and S5**). *MLO* clade II genes originally

730   evolved in ancient seed-producing plants, suggesting that genes from this clade could have sex-

731   related functions (Feechan et al., 2008; Jiwan et al., 2013; Zhou et al., 2013). On the other hand,

732   this male-specific duplication could represent a technical artefact caused by the fact that the

733   sequence of *CsMLO13* on chromosome Y was concatenated with its homologous version on

734   chromosome X, thus producing a false tandem duplication. To our knowledge, no studies have

735   documented in detail the structure and evolutionary relationships between orthologous *MLOs* in a

736   group of genomes from the same species. Plus, information on sex-specific distribution of *MLOs*

737   in comparable heterogametic sex plant systems is scarce, which makes the interpretation of this

738   finding difficult. This could ultimately indicate that clade II *MLOs* may not be solely related to

739   seed production or development, as suggested in other systems (Kusch et al., 2016). Recently, a

740   clade II *MLO* was shown to be dictating PM susceptibility in mungbean (Yundaeng et al., 2020).

741   However, apart from this example, only clade V *MLOs* have been shown to be involved in this

742   trait in dicots. In this study, one cannabis clade V *MLO*, *CsMLO1*, appears to be duplicated in three

743   different genomes (three out of the four THC-producing cultivars). If not an assembly artefact, the

744   presence of such an additional copy of a clade V *MLO* would make it tedious to obtain complete

745   immunity to PM. In other plants such as *Arabidopsis*, inactivation of all clade V *MLOs* is required

746   to achieve complete immunity, even though these genes unevenly contribute to susceptibility, with

747   *AtMLO2* playing a major role (Consonni et al., 2006). The presence of a single copy of *CsMLO1*

748   in the industrial hemp variety Finola and the CBD-dominant CBDRx (which ancestry has been

749   suggested to be 11% hemp, Grassa et al., 2021) could suggest that hemp is de facto less susceptible

750   or that attaining this target in hemp could be easier to achieve. On the other hand, both Jamaican

751   Lion cultivars investigated here have been described as being highly resistant to PM (McKernan

752   et al., 2020), even though both have this extra, apparently functional, *CsMLO1* copy. In this

753   particular case, resistance to PM is likely due to other genetic factors than a loss-of-susceptibility

754   that would have been obtained through deletion/mutation of clade V *MLOs* (see below).

755

756   **4.3   Upregulation of clade V MLOs triggered by powdery mildew infection**

757   The transcriptomic response of *Cannabis sativa* to PM performed here revealed that clade V

758   *CsMLOs* are rapidly triggered, at least six hours post-inoculation, upon infection by the pathogen

759   (**Figure 4**). Our results revealed that the two clade V *CsMLO* genes responded with different

760   activation patterns, potentially suggesting specific roles. As described above, all three clade V

761   *MLO* genes need to be inactivated in order to achieve complete immunity against PM in

762   *Arabidopsis* (Consonni et al., 2006). However, in other plants, not all members of clade V are *S*-

763   genes, and the inactivation of only a subset of the clade V *MLO* genes is required (Bai et al., 2008;

764   Pavan et al., 2011). In those cases, the precise identification of the exact genes involved in

765   susceptibility needs to rely on additional criteria. A shared element of all *MLO* genes involved in

766   PM susceptibility is that they respond to fungal penetration, showing significant upregulation as

767   soon as six hours after inoculation (Piffanelli et al., 2002; Bai et al., 2008; Zheng et al., 2013), and

768   candidate genes can thus be identified based on increased expression (Feechan et al., 2008; Pessina

769   et al., 2014). Indeed, being an *S*-gene, the expression of *MLO* is necessary for the successful

770   invasion of PM (Freialdenhoven et al., 1996; Lyngkjær et al., 2000; Zellerhoff et al., 2010). For

771   instance, in watermelon, upregulation was only observed for one clade V *MLO* (out of five),

772   *ClMLO12*, and only at time points corresponding to nine and 24 hours after inoculation with

773   *Podosphaera xanthii*, making it the prime candidate dictating PM susceptibility (Iovieno et al.,

774   2015). In apple, three genes (including two out of the four clade V *MLOs*, *MdMLO11* and

775   *MdMLO19*) were found to be significantly up-regulated after inoculation with PM, reaching about

776   2-fold compared to non-inoculated plants (Pessina et al., 2014). Similarly, three out of four

777   grapevine clade V *MLOs* (*VvMLO3*, *VvMLO4*, and *VvMLO17*) were induced during infection by

778   *Erysiphe necator* (Feechan et al., 2008). Interestingly, in both apple and grapevine, a clade VII

779   MLO was also found to respond to PM, but the significance of this finding is unclear. In our case,

780   both clade V *CsMLOs* (*CsMLO1* and *CsMLO4*) appeared responsive to PM infection, showing a

781   2-fold upregulation (FDR < 0.01). This was further supported by the presence of *cis*-acting

782   elements involved in gene overexpression by biotic and abiotic stresses in all of the clade V

783   *CsMLOs* characterized in this study. These combined results suggest that cannabis is in a similar

784   situation to that observed in *A. thaliana*, where all clade V genes could be involved in

785   susceptibility.

786

787   As of now, *VrMLO12* in mungbean (*Vigna radiata*), which clusters with clade II genes, is

788   the sole report of an *MLO* gene outside of clade V being clearly involved in PM susceptibility in

789   dicots (Yundaeng et al., 2020). In rice, a clade II *MLO* (*OsMLO3*) was also found to have an

790   expression pattern similar to clade V *MLOs* from *Arabidopsis* (Nguyen et al., 2016) and could

791   partially restore PM susceptibility in barley mutants, suggesting an involvement in plant defense

792   (Elliott et al., 2002). According to our results, clade II *CsMLOs* should not be considered

793   candidates, as none showed an induction following inoculation with *G. ambrosiae* emend.

794   (including *G. spadiceus*). However, other *CsMLOs* outside of clade V were found to be

795   differentially expressed, a situation not different from previous findings in apple (Pessina et al.,

796   2014), grapevine (Feechan et al., 2008), and tomato (Zheng et al., 2016). Interestingly, while only

797   one out of four clade V *MLO* showed pathogen-dependent upregulation in tomato (*SlMLO1*), there

798   is some overlap between non-clade V genes that respond to PM in tomato and cannabis. The

799   pathogen-triggered response is, however, not always in the same direction. In tomato, three clade

800   I *MLOs* (*SlMLO10*, *SlMLO13* and *SlMLO14*) were induced following an infection challenge with

801   *Oidium neolycopersici*. While the cannabis orthologs of the last two (*CsMLO9* and *CsMLO2*,

802   respectively) were also differentially expressed after inoculation with PM, their expression levels

803   rather decreased over time (**Supplementary Figure S2**). Similarly, the expression of the tomato

804   clade III *SlMLO4* significantly increased during infection, while that of its cannabis ortholog

805   *CsMLO7* decreased (**Supplementary Figure S2**). The sole tomato clade VI gene, *SlMLO16*, was

806   also found to be induced (Zheng et al., 2016). While its direct cannabis ortholog *CsMLO3* was not

807   found to be up-regulated, it was the case for the other cannabis clade VI gene, *CsMLO5*, for which

808   no ortholog exists in tomato or *Arabidopsis*. This particular gene was the only *CsMLOs* to be

809   induced following a challenge with *G. ambrosiae* emend. (including *G. spadiceus*) outside of clade

810   V.

811

812   Following our attempt to further characterize the two clade V CsMLOs and determine their

813   subcellular localization *in planta* using confocal laser scanning microscopy, we observed that

814   CsMLO4 was located in the plasma membrane while CsMLO1 was located in endomembrane-

815   associated compartments, including the plasma membrane, the endoplasmic reticulum and the

816   Golgi apparatus (**Figure 5**). In plants, the presence of a reticulate and network-looking pattern and

817   bright spots, as observed for CsMLO1, are typically associated with the endoplasmic reticulum

818   and Golgi stacks, respectively (Bassham et al., 2008). A time-series performed using confocal

819   microscopy demonstrated that the CsMLO1-EGFP fusion protein was extremely dynamic

820   compared to the CsMLO4-EGFP fusion protein (results not shown), supporting its implication in

821   intracellular trafficking through the endomembrane system. Many studies have demonstrated that

822   MLOs are localized in the plasma membrane (Devoto et al., 1999; Kim and Hwang, 2012; Nie et

823   al., 2015), supporting our observations with regards to CsMLO4. Other studies have indicated that

824   MLOs are associated with the plasma membrane and/or other endomembrane compartments, such

825   as the endoplasmic reticulum and the Golgi apparatus (Chen et al., 2009; Jones and Kessler, 2017;

826   Qin et al., 2019) and thus supporting our observations with regards to CsMLO1. To further support

827   our findings and to determine precise subcellular localization of CsMLOs, subcellular

828   fractionation studies as well as fluorescence colocalization with specific organelles markers should

829   be performed.

830

4.4   Genetic variation within clade V *CsMLOs*: a quest toward durable resistance

832   Several natural or induced loss-of-function mutations have been identified in *MLO* genes that

833   reduce susceptibility to PM (Büschges et al., 1997; Consonni et al., 2006; Pavan et al., 2011; Bai

834   et al., 2008; Wang et al., 2014). Barley *HvMlo1* is probably the most studied *MLO* gene, and most

835   mutations leading to loss-of-function (and thus, resistance to PM) in this gene appear to cluster in

836   the second and third cytoplasmic loops (Reinstädler et al., 2010). The functional importance of

837   those loops is still unclear, but evidence in other plants also points toward this particular region

838   (Fujimura et al., 2016; Acevedo-Garcia et al., 2017). Outside of this region, the integrity of

839   transmembrane domains, as well as certain invariant cysteine and proline residues, are critical for

840   the function and accumulation of *MLO* proteins (Elliott et al., 2005; Reinstädler et al., 2010).

841   Examination of polymorphisms among 32 cannabis cultivars (including one hemp variety)

842   identified only a single SNP in the second cytoplasmic loop of *CsMLO1*. No mutations were

843   identified among those invariant cysteine and proline residues, and no mutations were found either

844   in any of the strictly conserved residues or within transmembrane domains. This suggests that loss-

845   of-function mutations in *CsMLOs* could be rare or nonexistent among commercial cultivars,

846   complicating future breeding efforts. The potential lack of diversity among the cultivars included

847   in our analysis might also have impeded our ability to find causative loss-of-function mutations,

848   which could be a scenario likely generalizable to a significant part of the cannabis production

849   industry. Nevertheless, it is possible that mutations identified outside of those previously identified

850   regions could inactivate *CsMLOs*. Three mutations causing amino acid replacements were

851   identified in the first cytoplasmic loop of *CsMLO1*, and five mutations (four in *CsMLO1* and one

852   in *CsMLO4*) were identified in the first extracellular loop. The proximal part of the C-terminus of

853   MLO proteins contains a binding site for the cytoplasmic calcium sensor, calmodulin (Kim et al.,

854   2002a; Kim et al., 2002b). In barley MLO, binding of calmodulin to this domain appears to be

855   required for full susceptibility. Unfortunately, even though a high number of polymorphisms were

856   identified in the cytoplasmic C-terminus of both *CsMLO1* (six mutations) and *CsMLO4* (10

857   mutations), those mutations are not found within the calmodulin-binding domain but rather at the

858   distal end of the C-terminus. This might not be surprising, as this region is intrinsically disordered

859   (Kusch et al., 2016). Intrinsically disordered regions, i.e. regions lacking stable secondary

860   structures, usually exhibit greater amounts of non-synonymous mutations and other types of

861   polymorphisms because of the lack of structural constraints (Nilsson et al., 2011; Kusch et al.,

862   2016). Another study conducted on clade V *MLOs* had also revealed that both the first extracellular

863   loop and the C-terminus were under strong positive selection (Iovieno et al., 2015), which seems

864   in agreement with our observations.

865

866   The fact that potential loss-of-function mutations in clade V *MLOs* were not identified

867   among the 32 investigated genomes suggests that complete resistance to PM might be hard to find

868   among existing commercial cultivars. Nevertheless, such mutations might be present at a very low

869   frequency, especially in "wild" populations or in landraces that have infrequently been used in

870   breeding programs. For example, natural *mlo* alleles exist in barley but have only been found in

871   landraces from Ethiopia and Yemen (Reinstädler et al., 2010). Similarly, the natural loss-of-

872   function mutations in pea and tomato originated from wild accessions (Humphry et al., 2011; Bai

873   et al., 2008). This should reinforce the importance of preserving cannabis wild populations and

874   encourages efforts to establish germplasm repositories. However, *MLO*-based resistance to PM

875   being a recessive trait, and assuming that both *CsMLO1* and *CsMLO4* are involved, this would

876   mean that loss-of-function mutations would need to be bred as homozygous recessive for both

877   genes into elite plants (not considering that multiple copies of a given *CsMLO* might exist, as

878   suggested here for *CsMLO1*). In the absence of natural mutants, or to accelerate the

879   implementation of *mlo*-based resistance in breeding programs, induced mutagenesis and genome

880   editing might be interesting alternatives. While such approaches have not been optimized for

881   cannabis, loss-of-function mutations have been obtained through those in other crops, such as

882   barley (Reinstädler et al., 2010), wheat (Acevedo-Garcia et al., 2017; Wang et al., 2014), or tomato

883   (Nekrasov et al., 2017).

884

885   There might also be other routes to combating PM than MLO. In crops where gene-for-

886   gene interactions exist with PM (e.g., in cereals), a series of functional alleles confer complete

887   resistance against distinct sets of PM isolates (Bourras et al., 2019). Since a similar situation has

888   been observed between hop (*Humulus lupulus*, the closest relative of cannabis) and *Podosphaera*

889   *macularis* (Henning et al., 2011), it is likely that such gene-for-gene interactions exist between

890   cannabis and PM. In grapevine, resistance is usually considered polygenic, and there appear to be

891   a diverse range of responses to invasion by *Erysiphe necator*, from penetration resistance to the

892   induction of plant cell death (Feechan et al., 2011). Because cannabis has been excluded from

893   scientific research for decades, very little is known about the biology of PM infection (i.e. genetic

894   diversity, histology, host range), and no data on resistance levels among cultivars has been

895   published as of now. Still, a recent study observed higher copy numbers of a thaumatin-like protein

896   in several cannabis cultivars reported to be resistant to PM, and similar correlations were identified

897   with copy number variations of endochitinases and *CsMLOs* (McKernan et al., 2020). Thaumatin-

898   like proteins and endochitinases might thus represent additional targets potentially involved in

899   quantitative resistance to PM.

900

901   **5   Conclusion**

902

903   Genome-wide characterization of *CsMLOs* performed in this study indicated that genes from clade

904   V, which are immediately triggered upon infection by PM, are likely involved as early actors in

905   the fungal infection process by the plant. Polymorphism data generated for clade V *CsMLOs* in

906   multiple commercial cultivars suggested that loss-of-function mutations, required to achieve a

907   resistance phenotype, are rare events and potentially challenging to assemble, especially when

908   considering their recessive nature and the genetic redundancy of multiple clade V *CsMLOs*.

909   Preserving a diversified collection of feral and, ideally, landrace cannabis genetic backgrounds

910   while establishing coherent germplasm repositories could represent efficient strategies to compose

911   with this complex biological reality. This will allow the creation of novel and valuable knowledge

912   on the fundamental biology supporting the interaction between PM and cannabis, ultimately

913   leading to more sustainable horticultural and agro-industrial practices.

914

915

916 **Conflict of interest**

917 The authors declare that the research was conducted in the absence of any commercial or financial

918 relationships that could be construed as a potential conflict of interest.

919

920 **Author contributions**

921 NP, FOH and DLJ contributed to the conception and design of the study. NP performed the

922 experiments. NP, FOH and DLJ analyzed the data and drafted the manuscript. All authors revised

923 the manuscript, read, and approved the submitted version. DLJ contributed reagents, equipment,

924 and funds.

925

936

937 **Data availability statement**

938 All sequences generated in this study have been deposited in DDBJ/EMBL/GenBank under the

939 BioProject accession: PRJNA738505, SRA accessions: SRR14839036-50; and the BioProject

940 accession: PRJNA738519, SRA accessions: SRR14857079-110

941

942 **References**

943 Abbasi, W. A., Asif, A., Andleeb, S., and Minhas, F. U. A. A. (2017). CaMELS: *In silico*

944 prediction of calmodulin binding proteins and their binding sites. Proteins 85, 1724–1740.

945 doi:10.1002/prot.25330.

946

947   Ablazov, A., and Tombuloglu, H. (2016). Genome-wide identification of the mildew resistance

948   locus O (*MLO*) gene family in novel cereal model species *Brachypodium distachyon*. Eur. J. Plant

949   Pathol. 145, 239–253. doi:10.1007/s10658-015-0833-2.

950

951   AbuQamar, S., Moustafa, K., and Tran, L. S. (2017). Mechanisms and strategies of plant defense

952   against *Botrytis cinerea*. Crit. Rev. Biotechnol. 37, 262–274.

953   doi:10.1080/07388551.2016.1271767.

954

955   Acevedo-Garcia, J., Kusch, S., and Panstruga, R. (2014). Magical mystery tour: *MLO* proteins in

956   plant immunity and beyond. New Phytologist 204, 273–281. doi:10.1111/nph.12889.

957

958   Acevedo-Garcia, J., Spencer, D., Thieron, H., Reinstädler, A., Hammond-Kosack, K., Phillips, A.

959   L., et al. (2017). *mlo*-based powdery mildew resistance in hexaploid bread wheat generated by a

960   non-transgenic TILLING approach. Plant Biotechnol. J. 15, 367–378. doi:10.1111/pbi.12631.

961

962   Almagro Armenteros, J. J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson,

963   A., et al. (2019). Detecting sequence signals in targeting peptides using deep learning. Life Sci.

964   Alliance 2, e201900429. doi:10.26508/lsa.201900429.

965

966   Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., and Winther, O. (2017).

967   DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics 33,

968   3387–3395. doi:10.1093/bioinformatics/btx431.

969

970   Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak,

971   S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat.

972   Biotechnol. 37, 420–423. doi:10.1038/s41587-019-0036-z.

973

974   Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq-A Python framework to work with high-

975   throughput sequencing data. Bioinformatics 31, 166–169. doi:10.1093/bioinformatics/btu638.

976

977   Appiano, M., Pavan, S., Catalano, D., Zheng, Z., Bracuto, V., Lotti, C., et al. (2015). Identification
978   of candidate *MLO* powdery mildew susceptibility genes in cultivated Solanaceae and functional
979   characterization of tobacco *NtMLO1*. Transgenic Res. 24, 847–858. doi:10.1007/s11248-015-
980   9878-4.

981

982   Bai, Y., Pavan, S., Zheng, Z., Zappel, N. F., Reinstädler, A., Lotti, C., et al. (2008). Naturally
983   occurring broad-spectrum powdery mildew resistance in a Central American tomato accession is
984   caused by loss of *Mlo* function. Mol. Plant-Microbe Interact. 21, 30–39. doi:10.1094/MPMI-21-
985   1-0030.

986

987   Bassham, D. C., Brandizzi, F., Otegui, M. S., and Sanderfoot, A. A. (2008). The secretory system
988   of *Arabidopsis*. *Arabidopsis* Book 6, e0116. doi:10.1199/tab.0116.

989

990   Bidzinski, P., Noir, S., Shahi, S., Reinstädler, A., Gratkowska, D. M., and Panstruga, R. (2014).
991   Physiological characterization and genetic modifiers of aberrant root thigmomorphogenesis in
992   mutants of *Arabidopsis thaliana MILDEW LOCUS O* genes. Plant, Cell and Environ. 37, 2738–
993   2753. doi:10.1111/pce.12353.

994

995   Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., et al. (2021).
996   The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49, D344–
997   D354. doi:10.1093/nar/gkaa977.

998

999   Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina
1000   sequence data. Bioinformatics 30, 2114–2120. doi:10.1093/bioinformatics/btu170.

1001

1002   Bourras, S., Kunz, L., Xue, M., Praz, C. R., Müller, M. C., Kälin, C., et al. (2019). The *AvrPm3-
1003   Pm3* effector-NLR interactions control both race-specific resistance and host-specificity of cereal
1004   mildews on wheat. Nat. Commun. 10, 2292. doi:10.1038/s41467-019-10274-1.

1005

1006   Braun, U., Cook, R. T. A., Inman, A. J., and Shin, H. D. (2002). "The Taxonomy of the Powdery
1007   Mildew Fungi," in The Powdery Mildews: A Comprehensive Treatise, ed. R. R. Bélanger, W. R.
1008   Bushnell, A. J. Dik, and T. L. W. Carver (Saint Paul, MN, USA: APS Press), 13-55.

1009

1010   Braun, U., and Cook, R. T. A. (2012). Taxonomic Manual of the *Erysiphales* (Powdery Mildews).
1011   CBS Biodiversity Series No. 11. Utrecht, the Netherlands: CBS-KNAW Fungal Biodiversity
1012   Centre.

1013

1014   Briesemeister, S., Rahnenführer, J., and Kohlbacher, O. (2010). YLoc-an interpretable web server
1015   for predicting subcellular localization. Nucleic Acids Res. 38, 497–502. doi:10.1093/nar/gkq477.

1016

1017   Büschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., et al. (1997). The
1018   barley *Mlo* gene: A novel control element of plant pathogen resistance. Cell 88, 695–705.
1019   doi:10.1016/S0092-8674(00)81912-1.

1020

1021   Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009).
1022   BLAST+: Architecture and applications. BMC Bioinformatics 10, 421. doi:10.1186/1471-2105-
1023   10-421.

1024

1025   Chen, Y., Wang, Y., and Zhang, H. (2014). Genome-wide analysis of the mildew resistance locus
1026   o (*MLO*) gene family in tomato (*Solanum lycopersicum* L.). Plant Omics 7, 87–93.

1027

1028   Chen, Z., Hartmann, H. A., Wu, M.-J., Friedman, E. J., Chen, J.-G., Pulley, M., et al. (2006).
1029   Expression analysis of the *AtMLO* gene family encoding plant-specific seven-transmembrane
1030   domain proteins. Plant Mol. Biol. 60, 583–597. doi:10.1007/s11103-005-5082-x.

1031

1032   Chen, Z., Noir, S., Kwaaitaal, M., Hartmann, H. A., Wu, M. J., Mudgil, Y., et al. (2009). Two
1033   seven-transmembrane domain MILDEW RESISTANCE LOCUS O proteins cofunction in
1034   *Arabidopsis* root thigmomorphogenesis. Plant Cell 21, 1972–1991. doi:10.1105/tpc.108.062653.

1035

1036   Chiang, C., Layer, R. M., Faust, G. G., Lindberg, M. R., David, B., Garrison, E. P., et al. (2015).

1037   SpeedSeq: Ultra-fast personal genome analysis and interpretation. Nat. Methods 12, 966–968.

1038   doi:10.1038/nmeth.3505.

1039

1040   Chou, K. C., and Shen, H. B. (2010). Plant-mPLoc: A top-down strategy to augment the power for

1041   predicting plant protein subcellular localization. PLoS One 5, e11335.

1042   doi:10.1371/journal.pone.0011335.

1043

1044   Consonni, C., Humphry, M. E., Hartmann, H. A., Livaja, M., Durner, J., Westphal, L., et al. (2006).

1045   Conserved requirement for a plant host cell protein in powdery mildew pathogenesis. Nat. Genet.

1046   38, 716–720. doi:10.1038/ng1806.

1047

1048   Dangl, J. L., and Jones, J. D. G. (2001). Plant pathogens and integrated defence responses to

1049   infection. Nature 411, 826–833. doi:10.1038/35081161.

1050

1051   Debode, F., Janssen, E., and Berben, G. (2013). Development of 10 new screening PCR assays for

1052   GMO detection targeting promoters (pFMV, pNOS, pSSuAra, pTA29, pUbi, pRice actin) and

1053   terminators (t35S, tE9, tOCS, tg7). Eur. Food Res. Technol. 236, 659–669. doi:10.1007/s00217-

1054   013-1921-1.

1055

1056   Deshmukh, R., Singh, V. K., and Singh, B. D. (2014). Comparative phylogenetic analysis of

1057   genome-wide *Mlo* gene family members from *Glycine max* and *Arabidopsis thaliana*. Mol. Genet.

1058   Genomics 289, 345–359. doi:10.1007/s00438-014-0811-y.

1059

1060   Devoto, A., Hartmann, H. A., Piffanelli, P., Elliott, C., Simmons, C., Taramino, G., et al. (2003).

1061   Molecular phylogeny and evolution of the plant-specific seven-transmembrane *MLO* family. J.

1062   Mol. Evol. 56, 77–88. doi:10.1007/s00239-002-2382-5.

1063

1064   Devoto, A., Piffanelli, P., Nilsson, I. M., Wallin, E., Panstruga, R., von Heijne, G., et al. (1999).

1065   Topology, subcellular localization, and sequence diversity of the *Mlo* family in plants. J. Biol.

1066   Chem. 274, 34993–35004. doi:10.1074/jbc.274.49.34993.

1067

1068   Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR:

1069   Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

1070   doi:10.1093/bioinformatics/bts635.

1071

1072   Dobson, L., Reményi, I., and Tusnády, G. E. (2015). CCTOP: A Consensus Constrained TOPology

1073   prediction web server. Nucleic Acids Res. 43, W408–W412. doi:10.1093/nar/gkv451.

1074

1075   Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB

1076   transcription factors in *Arabidopsis*. Trends Plant Sci. 15, 573–581.

1077   doi:10.1016/j.tplants.2010.06.005.

1078

1079   Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high

1080   throughput. Nucleic Acids Res. 32, 1792–1797. doi:10.1093/nar/gkh340.

1081

1082   Elliott, C., Müller, J., Miklis, M., Bhat, R. A., Schulze-Lefert, P., and Panstruga, R. (2005).

1083   Conserved extracellular cysteine residues and cytoplasmic loop-loop interplay are required for

1084   functionality of the heptahelical *MLO* protein. Biochem. J. 385, 243–254.

1085   doi:10.1042/BJ20040993.

1086

1087   Elliott, C., Zhou, F., Spielmeyer, W., Panstruga, R., and Schulze-Lefert, P. (2002). Functional

1088   conservation of wheat and rice *Mlo* orthologs in defense modulation to the powdery mildew

1089   fungus. Mol. Plant-Microbe Interact. 15, 1069–1077. doi:10.1094/MPMI.2002.15.10.1069.

1090

1091   Fan, C., Chen, Y., and Long, M. (2008). Recurrent tandem gene duplication gave rise to

1092   functionally divergent genes in *Drosophila*. Mol. Biol. Evol. 25, 1451–1458.

1093   doi:10.1093/molbev/msn089.

1094

1095   Farinas, C., and Peduto Hand, F. (2020). First Report of *Golovinomyces spadiceus* Causing

1096   Powdery Mildew on Industrial Hemp (*Cannabis sativa*) in Ohio. Plant Dis. 104, 2727.

1097   doi:10.1094/PDIS-01-20-0198-PDN.

1098

1099 Feechan, A., Jermakow, A. M., Torregrosa, L., Panstruga, R., and Dry, I. B. (2008). Identification
1100 of grapevine *MLO* gene candidates involved in susceptibility to powdery mildew. Funct. Plant
1101 Biol. 35, 1255–1266. doi:10.1071/FP08173.

1102

1103 Feechan, A., Kabbara, S., and Dry, I. B. (2011). Mechanisms of powdery mildew resistance in the
1104 Vitaceae family. Mol. Plant Pathol. 12, 263–274. doi:10.1111/j.1364-3703.2010.00668.x.

1105

1106 Freialdenhoven, A., Peterhänsel, C., Kurth, J., Kreuzaler, F., and Schulze-Lefert, P. (1996).
1107 Identification of genes required for the function of non-race-specific *mlo* resistance to powdery
1108 mildew in barley. Plant Cell 8, 5–14. doi:10.1105/tpc.8.1.5.

1109

1110 Freisleben, R., and Lein, A. (1942). Über die Auffindung einer mehltauresistenten Mutante nach
1111 Röntgenbestrahlung einer anfälligen reinen Linie von Sommergerste. Naturwissenschaften 30,
1112 608. doi:10.1007/BF01488231.

1113

1114 Fujimura, T., Sato, S., Tajima, T., and Arai, M. (2016). Powdery mildew resistance in the Japanese
1115 domestic tobacco cultivar Kokubu is associated with aberrant splicing of *MLO* orthologues. Plant
1116 Pathol. 65, 1358–1365. doi:10.1111/ppa.12498.

1117

1118 Glawe, D. A. (2008). The powdery mildews: A review of the world's most familiar (yet poorly
1119 known) plant pathogens. Annu. Rev. Phytopathol. 46, 27–51.
1120 doi:10.1146/annurev.phyto.46.081407.104740.

1121

1122 Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Timothy Motley, S., et
1123 al. (2021). A new *Cannabis* genome assembly associates elevated cannabidiol (CBD) with hemp
1124 introgressed into marijuana. New Phytol. 230, 1665–1679. doi:10.1111/nph.17243.

1125

1126 Guigó, R., Agarwal, P., Abril, J. F., Burset, M., and Fickett, J. W. (2000). An assessment of gene
1127 prediction accuracy in large DNA sequences. Genome Res. 10, 1631–1642.
1128 doi:10.1101/gr.122800.

1129

1130   Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., et al. (2020). *RIdeogram*: Drawing SVG

1131   graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput. Sci. 6, e251.

1132   doi:10.7717/peerj-cs.251.

1133

1134   Hartley, J. L., Temple, G. F., and Brasch, M. A. (2000). DNA cloning using *in vitro* site-specific

1135   recombination. Genome Res. 10, 1788–1795. doi:10.1101/gr.143000.

1136

1137   Health Canada. (2020). Licensed cultivators, processors and sellers of cannabis under the Cannabis

1138   Act.   https://www.canada.ca/en/health-canada/services/drugs-medication/cannabis/industry-

1139   licensees-applicants/licensed-cultivators-processors-sellers.html [Accessed December 1, 2020].

1140

1141   Henning, J. A., Townsend, M. S., Gent, D. H., Bassil, N., Matthews, P., Buck, E., et al. (2011).

1142   QTL mapping of powdery mildew susceptibility in hop (*Humulus lupulus* L.). Euphytica 180, 411–

1143   420. doi:10.1007/s10681-011-0403-4.

1144

1145   Hilbert, M., Novero, M., Rovenich, H., Mari, S., Grimm, C., Bonfante, P., et al. (2020). *MLO*

1146   differentially regulates barley root colonization by beneficial endophytic and mycorrhizal fungi.

1147   Front. Plant Sci. 10, 1–11. doi:10.3389/fpls.2019.01678.

1148

1149   Holt, C., and Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database

1150   management tool for second-generation genome projects. BMC Bioinformatics 12, 491.

1151   doi:10.1186/1471-2105-12-491.

1152

1153   Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic

1154   trees. Bioinformatics 17, 754–755. doi:10.1093/bioinformatics/17.8.754.

1155

1156   Humphry, M., Reinstädler, A., Ivanov, S., Bisseling, T., and Panstruga, R. (2011). Durable broad-

1157   spectrum powdery mildew resistance in pea *er1* plants is conferred by natural loss-of-function

1158   mutations in *PsMLO1*. Mol. Plant Pathol. 12, 866–878. doi:10.1111/j.1364-3703.2011.00718.x.

1159

1160 Ingvardsen, C. R., Massange-Sánchez, J. A., Borum, F., Uauy, C., and Gregersen, P. L. (2019).

1161 Development of *mlo*-based resistance in tetraploid wheat against wheat powdery mildew. Theor.

1162 Appl. Genet. 132, 3009–3022. doi:10.1007/s00122-019-03402-4.

1163

1164 Iovieno, P., Bracuto, V., Pavan, S., Lotti, C., Ricciardi, L., and Andolfo, G. (2016). Identification

1165 and functional inference on the *Mlo*-family in Viridiplantae. J. Plant Pathol. 98, 587–594.

1166 doi:10.4454/JPP.V98I3.027.

1167

1168 Iovieno, P., Andolfo, G., Schiavulli, A., Catalano, D., Ricciardi, L., Frusciante, L., et al. (2015).

1169 Structure, evolution and functional inference on the *Mildew Locus O* (*MLO*) gene family in three

1170 cultivated Cucurbitaceae *spp*. BMC Genomics 16, 1112. doi:10.1186/s12864-015-2325-3.

1171

1172 Jacott, C. N., Charpentier, M., Murray, J. D., and Ridout, C. J. (2020). *Mildew Locus O* facilitates

1173 colonization by arbuscular mycorrhizal fungi in angiosperms. New Phytol. 227, 343–351.

1174 doi:10.1111/nph.16465.

1175

1176 Jiang, K., and Goertzen, L. R. (2011). Spliceosomal intron size expansion in domesticated

1177 grapevine (*Vitis vinifera*). BMC Res. Notes 4, 52. doi:10.1186/1756-0500-4-52.

1178

1179 Jiwan, D., Roalson, E. H., Main, D., and Dhingra, A. (2013). Antisense expression of peach

1180 mildew resistance locus O (*PpMlo1*) gene confers cross-species resistance to powdery mildew in

1181 *Fragaria x ananassa*. Transgenic Res. 22, 1119–1131. doi:10.1007/s11248-013-9715-6.

1182

1183 Jones, D. S., and Kessler, S. A. (2017). Cell type-dependent localization of *MLO* proteins. Plant

1184 Signal. Behav. 12, e1393135. doi:10.1080/15592324.2017.1393135.

1185

1186 Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan

1187 5: Genome-scale protein function classification. Bioinformatics 30, 1236–1240.

1188 doi:10.1093/bioinformatics/btu031.

1189

1190   Jørgensen, I. H. (1992). Discovery, characterization and exploitation of *Mlo* powdery mildew
1191   resistance in barley. Euphytica 63, 141–152. doi:10.1007/BF00023919.

1193   Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY™ vectors for *Agrobacterium*-
1194   mediated plant transformation. Trends Plant Sci. 7, 193–195. doi:10.1016/s1360-1385(02)02251-
1195   3.

1197   Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:
1198   Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
1199   doi:10.1093/molbev/mst010.

1201   Kaufmann, H., Qiu, X., Wehmeyer, J., and Debener, T. (2012). Isolation, molecular
1202   characterization, and mapping of four rose *MLO* orthologs. Front. Plant Sci. 3, 244.
1203   doi:10.3389/fpls.2012.00244.

1205   Kessler, S. A., Shimosato-Asano, H., Keinath, N. F., Wuest, S. E., Ingram, G., Panstruga, R., et
1206   al. (2010). Conserved molecular components for pollen tube reception and fungal invasion.
1207   Science 330, 968–971. doi:10.1126/science.1195211.

1209   Kim, D. S., and Hwang, B. K. (2012). The pepper *MLO* gene, *CaMLO2*, is involved in the
1210   susceptibility cell-death response and bacterial and oomycete proliferation. Plant J. 72, 843–855.
1211   doi:10.1111/tpj.12003.

1213   Kim, M. C., Panstruga, R., Elliott, C., Möller, J., Devoto, A., Yoon, H. W., et al. (2002).
1214   Calmodulin interacts with *MLO* protein to regulate defence against mildew in barley. Nature 416,
1215   447–450. doi:10.1038/416447a.

1217   Kim, M. C., Lee, S. H., Kim, J. K., Chun, H. J., Choi, M. S., Chung, W. S., et al. (2002). *Mlo*, a
1218   modulator of plant defense and cell death, is a novel calmodulin-binding protein: Isolation and
1219   characterization of a rice *Mlo* homologue. J. Biol. Chem. 277, 19304–19314.
1220   doi:10.1074/jbc.M108478200.

1221

1222 Konishi, S., Sasakuma, T., and Sasanuma, T. (2010). Identification of novel *Mlo* family members
1223 in wheat and their genetic characterization. Genes Genet. Syst. 85, 167–175.
1224 doi:10.1266/ggs.85.167.

1225

1226 Kusch, S., Pesch, L., and Panstruga, R. (2016). Comprehensive phylogenetic analysis sheds light
1227 on the diversity and origin of the *MLO* family of integral membrane proteins. Genome Biol. Evol.
1228 8, 878–895. doi:10.1093/gbe/evw036.

1229

1230 Lallemand, T., Leduc, M., Landès, C., Rizzon, C., and Lerat, E. (2020). An overview of duplicated
1231 gene detection methods: Why the duplication mechanism has to be accounted for in their choice.
1232 Genes 11, 1046. doi:10.3390/genes11091046.

1233

1234 Laverty, K. U., Stout, J. M., Sullivan, M. J., Shah, H., Gill, N., Holbrook, L., et al. (2019). A
1235 physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the *THC*/*CBD*
1236 *acid synthase* loci. Genome Res. 29, 146–156. doi:10.1101/gr.242594.118.

1237

1238 Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). Voom: Precision weights unlock linear
1239 model analysis tools for RNA-seq read counts. Genome Biol. 15, R29. doi:10.1186/gb-2014-15-
1240 2-r29.

1241

1242 Lefort, V., Longueville, J. E., and Gascuel, O. (2017). SMS: Smart Model Selection in PhyML.
1243 Mol. Biol. Evol. 34, 2422–2424. doi:10.1093/molbev/msx149.

1244

1245 Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., et al.
1246 (2019). NGPhylogeny.fr: New generation phylogenetic services for non-specialists. Nucleic Acids
1247 Res. 47, W260–W265. doi:10.1093/nar/gkz303.

1248

1249 Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van De Peer, Y., et al. (2002).
1250 PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico*
1251 analysis of promoter sequences. Nucleic Acids Res. 30, 325–327. doi:10.1093/nar/30.1.325.

1252

1253   Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new

1254   developments. Nucleic Acids Res. 47, W256–W259. doi:10.1093/nar/gkz239.

1255

1256   Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping

1257   and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–

1258   2993. doi:10.1093/bioinformatics/btr509.

1259

1260   Liu, Q., and Zhu, H. (2008). Molecular evolution of the *MLO* gene family in *Oryza sativa* and

1261   their functional divergence. Gene 409, 1–10. doi:10.1016/j.gene.2007.10.031.

1262

1263   Lyngkjær, M. F., and Carver, T. L. W. (2000). Conditioning of cellular defence responses to

1264   powdery mildew in cereal leaves by prior attack. Mol. Plant Pathol. 1, 41–49. doi:10.1046/j.1364-

1265   3703.2000.00006.x.

1266

1267   Lyngkjær, M. F., Newton, A. C., Atzema, J. L., and Baker, S. J. (2000). The barley *mlo*-gene: an

1268   important powdery mildew resistance source. Agronomie 20, 745–756.

1269   doi:10.1051/agro:2000173.

1270

1271   McKernan, K. J., Helbert, Y., Kane, L. T., Ebling, H., Zhang, L., Liu, B., et al. (2018).

1272   Cryptocurrencies and Zero Mode Wave guides: An unclouded path to a more contiguous *Cannabis*

1273   *sativa* L. genome assembly. [Preprint]. Available at: https://osf.io/7d968/ (Accessed June 3, 2021).

1274

1275   McKernan, K., Helbert, Y., Kane, L., Ebling, H., Zhang, L., Liu, B., et al. (2020). Sequence and

1276   annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid

1277   synthesis   and   pathogen   resistance   genes   [Preprint].   Available   at:

1278   https://www.biorxiv.org/content/10.1101/2020.01.03.894428v1 (Accessed June 3, 2021).

1279

1280   Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et

1281   al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Res. 49, D412–D419.

1282   doi:10.1093/nar/gkaa913.

1283

1284   Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D., and Kamoun, S. (2017). Rapid generation
1285   of a transgene-free powdery mildew resistant tomato by genome deletion. Sci. Rep. 7, 482.
1286   doi:10.1038/s41598-017-00578-x.

1287

1288   Nelson, C. W., Moncla, L. H., and Hughes, A. L. (2015). SNPGenie: Estimating evolutionary
1289   parameters to detect natural selection using pooled next-generation sequencing data.
1290   Bioinformatics 31, 3709–3711. doi:10.1093/bioinformatics/btv449.

1291

1292   Nguyen, V. N. T., Vo, K. T. X., Park, H., Jeon, J. S., and Jung, K. H. (2016). A systematic view
1293   of the *MLO* family in rice suggests their novel roles in morphological development, diurnal
1294   responses, the light-signaling pathway, and various stress responses. Front. Plant Sci. 7, 1413.
1295   doi:10.3389/fpls.2016.01413.

1296

1297   Nie, J., Wang, Y., He, H., Guo, C., Zhu, W., Pan, J., et al. (2015). Loss-of-function mutations in
1298   *CsMLO1* confer durable powdery mildew resistance in cucumber (*Cucumis sativus* L.). Front.
1299   Plant Sci. 6, 1155. doi:10.3389/fpls.2015.01155.

1300

1301   Nilsson, J., Grahn, M., and Wright, A. P. H. (2011). Proteome-wide evidence for enhanced positive
1302   Darwinian selection within intrinsically disordered regions in proteins. Genome Biol. 12, R65.
1303   doi:10.1186/gb-2011-12-7-r65.

1304

1305   Pavan, S., Schiavulli, A., Appiano, M., Marcotrigiano, A. R., Cillo, F., Visser, R. G. F., et al.
1306   (2011). Pea powdery mildew *er1* resistance is associated to loss-of-function mutations at a *MLO*
1307   homologous locus. Theor. Appl. Genet. 123, 1425–1431. doi:10.1007/s00122-011-1677-6.

1308

1309   Pépin, N., Punja, Z. K., and Joly, D. L. (2018). Occurrence of Powdery Mildew Caused by
1310   *Golovinomyces cichoracearum sensu lato* on *Cannabis sativa* in Canada. Plant Dis. 102, 2644.
1311   doi:10.1094/PDIS-04-18-0586-PDN.

1312

1313   Pessina, S., Lenzi, L., Perazzolli, M., Campa, M., Dalla Costa, L., Urso, S., et al. (2016).
1314   Knockdown of *MLO* genes reduces susceptibility to powdery mildew in grapevine. Hortic. Res. 3,
1315   16016. doi:10.1038/hortres.2016.16.

1316

1317   Pessina, S., Pavan, S., Catalano, D., Gallotta, A., Visser, R. G. F., Bai, Y., et al. (2014).
1318   Characterization of the *MLO* gene family in Rosaceae and gene expression analysis in *Malus*
1319   *domestica*. BMC Genomics 15, 618. doi:10.1186/1471-2164-15-618.

1320

1321   Piffanelli, P., Zhou, F., Casais, C., Orme, J., Jarosch, B., Schaffrath, U., et al. (2002). The barley
1322   *MLO* modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. Plant
1323   Physiol. 129, 1076–1085. doi:10.1104/pp.010954.

1324

1325   Pilkington, S. M., Crowhurst, R., Hilario, E., Nardozza, S., Fraser, L., Peng, Y., et al. (2018). A
1326   manually annotated *Actinidia chinensis* var. *chinensis* (kiwifruit) genome highlights the challenges
1327   associated with draft genomes and gene prediction in plants. BMC Genomics 19, 257.
1328   doi:10.1186/s12864-018-4656-3.

1329

1330   Polanco, C., Sáenz de Miera, L. E., Bett, K., and de la Vega, M. P. (2018). A genome-wide
1331   identification and comparative analysis of the lentil *MLO* genes. PLoS One 13, e0194945.
1332   doi:10.1371/journal.pone.0194945.

1333

1334   Punja, Z. K. (2021). Emerging diseases of *Cannabis sativa* and sustainable management. Pest
1335   Manag. Sci. doi:10.1002/ps.6307.

1336

1337   Punja, Z. K., Collyer, D., Scott, C., Lung, S., Holmes, J., and Sutton, D. (2019). Pathogens and
1338   molds affecting production and quality of *Cannabis sativa* L. Front. Plant Sci. 10, 1120.
1339   doi:10.3389/fpls.2019.01120.

1340

1341   Python Software Foundation. (2020). Python Language Reference, version 3.7.3. Available at:
1342   http://www.python.org.

1343

1344 Qin, B., Wang, M., He, H.-X., Xiao, H.-X., Zhang, Y., and Wang, L.-F. (2019). Identification and

1345 characterization of a potential candidate *MLO* gene conferring susceptibility to powdery mildew

1346 in rubber tree. Phytopathology 109, 1236–1245. doi:10.1094/PHYTO-05-18-0171-R.

1347

1348 Reinstädler, A., Müller, J., Czembor, J. H., Piffanelli, P., and Panstruga, R. (2010). Novel induced

1349 *mlo* mutant alleles in combination with site-directed mutagenesis reveal functionally important

1350 domains in the heptahelical barley *Mlo* protein. BMC Plant Biol. 10, 31. doi:10.1186/1471-2229-

1351 10-31.

1352

1353 Rispail, N., and Rubiales, D. (2016). Genome-wide identification and comparison of legume *MLO*

1354 gene family. Sci. Rep. 6, 32673. doi:10.1038/srep32673.

1355

1356 Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: A Bioconductor package for

1357 differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

1358 doi:10.1093/bioinformatics/btp616.

1359

1360 RStudio Team (2020). RStudio: Integrated development for R. Available at: https://rstudio.com.

1361

1362 Seoighe, C., and Korir, P. K. (2011). Evidence for intron length conservation in a set of mammalian

1363 genes associated with embryonic development. BMC Bioinformatics 12(Suppl 9), S16.

1364 doi:10.1186/1471-2105-12-S9-S16.

1365

1366 Shen, Q., Zhao, J., Du, C., Xiang, Y., Cao, J., and Qin, X. (2012). Genome-scale identification of

1367 *MLO* domain-containing genes in soybean (*Glycine max* L. Merr.). Genes Genet. Syst. 87, 89–98.

1368 doi:10.1266/ggs.87.89.

1369

1370 Shi, J., Wan, H., Zai, W., Xiong, Z., and Wu, W. (2020). Phylogenetic relationship of plant *MLO*

1371 genes and transcriptional response of *MLO* genes to *Ralstonia solanacearum* in tomato. Genes 11,

1372 487. doi:10.3390/genes11050487.

1373

1374 Singh, V. K., Singh, A. K., Chand, R., and Singh, B. D. (2012). Genome wide analysis of disease
1375 resistance *MLO* gene family in sorghum [*Sorghum bicolor* L. Moench]. Journal of Plant Genomics
1376 2, 18–27.

1378 Small, E. (2015). Evolution and Classification of *Cannabis sativa* (Marijuana, Hemp) in Relation
1379 to Human Utilization. Bot. Rev. 81, 189–294. doi:10.1007/s12229-015-9157-3.

1381 Statistics Canada. (2021). Table 20-10-0008-01 Retail trade sales by province and territory (x
1382 1,000). doi:10.25318/2010000801-eng [Accessed January 21, 2021].

1384 Szarka, D., Tymon, L., Amsden, B., Dixon, E., Judy, J., and Gauthier, N. (2019). First Report of
1385 Powdery Mildew Caused by *Golovinomyces spadiceus* on Industrial Hemp (*Cannabis sativa*) in
1386 Kentucky. Plant Dis. 103, 1773. doi:10.1094/PDIS-01-19-0049-PDN.

1388 Thompson, G. R., Tuscano, J. M., Dennis, M., Singapuri, A., Libertini, S., Gaudino, R., et al.
1389 (2017). A microbiome assessment of medical marijuana. Clin. Microbiol. Infect. 23, 269–270.
1390 doi:10.1016/j.cmi.2016.12.001.

1392 Tørresen, O. K., Star, B., Mier, P., Andrade-Navarro, M. A., Bateman, A., Jarnot, P., et al. (2019).
1393 Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome
1394 and protein databases. Nucleic Acids Res. 47, 10994–11006. doi:10.1093/nar/gkz841.

1396 Treangen, T. J., and Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing:
1397 computational challenges and solutions. Nat. Rev. Genet. 13, 36–46. doi:10.1038/nrg3117.

1399 Tully, J. P., Hill, A. E., Ahmed, H. M. R., Whitley, R., Skjellum, A., and Mukhtar, M. S. (2014).
1400 Expression-based network biology identifies immune-related functional modules involved in plant
1401 defense. BMC Genomics 15, 421. doi:10.1186/1471-2164-15-421.

1403 Van Rossum, G., and Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA:
1404 CreateSpace.

1405

1406   Van Schie, C. C. N., and Takken, F. L. W. (2014). Susceptibility genes 101: How to be a good

1407   host. Annu. Rev. Phytopathol. 52, 551–581. doi:10.1146/annurev-phyto-102313-045854.

1408

1409   Vielba-Fernández, A., Polonio, Á., Ruiz-Jiménez, L., de Vicente, A., Pérez-García, A., and

1410   Fernández-Ortuño, D. (2020). Fungicide resistance in powdery mildew fungi. Microorganisms 8,

1411   1–34. doi:10.3390/microorganisms8091431.

1412

1413   Wan, D. Y., Guo, Y., Cheng, Y., Hu, Y., Xiao, S., Wang, Y., et al. (2020). CRISPR/Cas9-mediated

1414   mutagenesis of *VvMLO3* results in enhanced resistance to powdery mildew in grapevine (*Vitis*

1415   *vinifera*). Hortic. Res. 7, 116. doi:10.1038/s41438-020-0339-8.

1416

1417   Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of

1418   three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew.

1419   Nat. Biotechnol. 32, 947–951. doi:10.1038/nbt.2969.

1420

1421   Weldon, W. A., Ullrich, M. R., Smart, L. B., Smart, C. D., and Gadoury, D. M. (2020). Cross-

1422   Infectivity of Powdery Mildew Isolates Originating from Hemp (*Cannabis sativa*) and Japanese

1423   Hop (*Humulus japonicus*) in New York. Plant Heal. Prog. 21, 47–53. doi:10.1094/PHP-09-19-

1424   0067-RS.

1425

1426   Whibley, A., Kelley, J. L., and Narum, S. R. (2021). The changing face of genome assemblies:

1427   Guidance on achieving high-quality reference genomes. Mol. Ecol. Resour. 21, 641–652.

1428   doi:10.1111/1755-0998.13312.

1429

1430   Win, K. T., Zhang, C., and Lee, S. (2018). Genome-wide identification and description of *MLO*

1431   family genes in pumpkin (*Cucurbita maxima* Duch.). Hortic. Environ. Biotechnol. 59, 397–410.

1432   doi:10.1007/s13580-018-0036-9.

1433

1434   Wiseman, M. S., Bates, T., Garfinkel, A., Ocamb, C. M., and Gent, D. H. (2021). First report of

1435   Powdery Mildew Caused by *Golovinomyces ambrosiae* on *Cannabis sativa* in Oregon. Plant Dis.

1436   doi:10.1094/PDIS-11-20-2455-PDN.

1437

1438   Wu, J. Y., Xiao, J. F., Wang, L. P., Zhong, J., Yin, H. Y., Wu, S. X., et al. (2013). Systematic

1439   analysis of intron size and abundance parameters in diverse lineages. Sci. China Life Sci. 56, 968–

1440   974. doi:10.1007/s11427-013-4540-y.

1441

1442   Yandell, M., and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. Nat. Rev.

1443   Genet. 13, 329–342. doi:10.1038/nrg3174.

1444

1445   Yi, J., An, S., and An, G. (2014). *OsMLO12*, encoding seven transmembrane proteins, is involved

1446   with pollen hydration in rice. Plant Reprod. 27, 169–180. doi:10.1007/s00497-014-0249-8.

1447

1448   Yundaeng, C., Somta, P., Chen, J., Yuan, X., Chankaew, S., Srinives, P., et al. (2020). Candidate

1449   gene mapping reveals *VrMLO12* (*MLO* Clade II) is associated with powdery mildew resistance in

1450   mungbean (*Vigna radiata* [L.] Wilczek). Plant Sci. 298, 110594.

1451   doi:10.1016/j.plantsci.2020.110594.

1452

1453   Zellerhoff, N., Himmelbach, A., Dong, W., Bieri, S., Schaffrath, U., and Schweizer, P. (2010).

1454   Nonhost resistance of barley to different fungal pathogens is associated with largely distinct,

1455   quantitative transcriptional responses. Plant Physiol. 152, 2053–2066.

1456   doi:10.1104/pp.109.151829.

1457

1458   Zheng, Z., Appiano, M., Pavan, S., Bracuto, V., Ricciardi, L., Visser, R. G. F., et al. (2016).

1459   Genome-wide study of the tomato *SlMLO* gene family and its functional characterization in

1460   response to the powdery mildew fungus *Oidium neolycopersici*. Front. Plant Sci. 7, 380.

1461   doi:10.3389/fpls.2016.00380.

1462

1463   Zheng, Z., Nonomura, T., Appiano, M., Pavan, S., Matsuda, Y., Toyoda, H., et al. (2013). Loss of

1464   function in *Mlo* orthologs reduces susceptibility of pepper and tomato to powdery mildew disease

1465   caused by *Leveillula taurica*. PLoS One 8, e70723. doi:10.1371/journal.pone.0070723.

1466

1467   Zhou, S. J., Jing, Z., and Shi, J. L. (2013). Genome-wide identification, characterization, and

1468   expression analysis of the *MLO* gene family in *Cucumis sativus*. Genet. Mol. Res. 12, 6565–6578.

1469   doi:10.4238/2013.December.11.8.

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

1471 [a] *CsMLO15-CBDRx* is not predicted in the CBDRx genome, as compared to the other four genomes analyzed.

1472 [b] Number of transmembrane domains (TMs) in the predicted protein, as determined by the CCTOP online prediction server (Dobson et al., 2015).

1473 [c] Subcellular localization as predicted by three online prediction servers: Plant-mPLoc (Chou and Shen, 2010), YLoc-HighRes (Briesmeister et al., 2010), and DeepLoc 1.0 (Armenteros et al., 2017).

1474 [d] Number of conserved amino acids out of the 30 identified in Elliott et al., 2005.

1475 [e] Number of conserved amino acids out of the 58 identified in Kusch et al., 2016.

1476 **Figure 1. Physical map of *MLO* genes identified in the chromosome-level assemblies of CBDRx,**

1477 **Finola and Purple Kush**. Each thick white bar represents a chromosome, with its size exactly proportional

1478 to the actual size in base pairs. Numbers below the chromosomes represent the "unified" chromosome

1479 numbers according to the CBDRx reference and the orthology analysis that we performed on all genomes.

1480 Colors indicate MLO clades from I to VII. In each CsMLO clade, each gene is designated by a different

1481 symbol (circle, square or triangle). Homology relationships between the three genomes are displayed with

1482 solid lines of colors corresponding to each MLO clade. The *MLO* gene from clade IV (herein named

1483 *CsMLO15*) is absent from the CBDRx assembly. Chromosomes 5 and 9 do not carry any *CsMLO* gene and

1484 are thus not represented in this figure.

1485

1486 **Figure 2. Intron-exon organization of the 14 manually curated *CsMLO* genes identified in the CBDRx**

1487 **genome.** Exons are shown as rectangles and introns as lines. The exon color code simply allows

1488 demonstrating exon conservation across all sequences. The numbers above exons indicate the exon's length

1489 (bp). Note that the STOP codon (3 bp) is included in the last exon's length. Sequences exhibiting one or

1490 several large introns are severed where necessary (i.e. *CsMLO4*, *CsMLO5*, *CsMLO6* and *CsMLO10*) and

1491 are therefore not drawn to scale.

1492

1493 **Figure 3. Phylogenetic relationships of CsMLOs based on Maximum Likelihood analysis.**

1494 Phylogenetic tree of manually curated CsMLO proteins (bold) with MLO proteins from selected species

1495 (*Arabidopsis thaliana*, *Prunus persica*, *Vitis vinifera*, *Hordeum vulgare* and *Zea mays*). *Chlamydomonas*

1496 *reinhardtii* was used as an outgroup. Phylogenetic relationships were estimated using the Maximum

1497 Likelihood method implemented in PhyML+SMS with 1,000 bootstrap independent replicates. The seven

1498 defined clades are indicated, as well as potential subclades identified in this study (inner circles). Number

1499 on a node indicates the percentage of bootstrap when higher than 65% (black), or the posterior probabilities

1500 of major clades and subclades, according to a Bayesian phylogenetic inference performed on the same

1501 alignment (red) (**Supplementary Figure S1**). MLOs with one asterisk (*) have been experimentally

1502 demonstrated to be required for PM susceptibility (Büschges et al., 1997; Feechan et al., 2008; Wan et al.,

1503 2020), while MLOs with two asterisks (**) have been identified as main probable candidates for PM

1504 susceptibility (Pessina et al., 2016).

1505

1506 **Figure 4**. **Transcriptomic response of clade V *CsMLO* genes following inoculations with powdery**

1507 **mildew**. Time series analysis of the infection of *Cannabis sativa* leaves by PM, showing average gene

1508 expression for *CsMLO1* (blue triangles) and *CsMLO4* (blue diamonds). Gene expression is displayed on

1509 the y-axis as the average logarithmic value of the Counts Per Million (log2(CPM)) at each time point

1510   (displayed on the x-axis, n = 3 per time point). Time points: no infection / control (T0), 6 hours post-

1511   inoculation (6H), 24 hours post-inoculation (1D), 3 days post-inoculation (3D), and 8 days post-inoculation

1512   (8D).

1513

1514   **Figure 5. Subcellular localization of *CsMLO1* and *CsMLO4* as observed by confocal laser scanning**

1515   **microscopy.** Transient expression of *35S::CsMLO1-EGFP* and *35S::CsMLO4-EGFP* constructs in

1516   *Nicotiana benthamiana* leaf epidermal cells shown 3 days after agro-infiltration. Scale bars = 25 μm.

1517

1518   **Figure 6. Protein structure and polymorphism analysis of the two clade V genes (*CsMLO1* and**

1519   ***CsMLO4*) in 32 distinct cannabis cultivars aligned against the CBDRx reference genome.** Panels A

1520   and B exhibit all non-synonymous nucleic acid substitutions (A) and amino acid replacements (B) identified

1521   in *CsMLO1* (located on chromosome 1) from 32 hemp and drug-type cannabis cultivars. Panels C and D

1522   exhibit all non-synonymous nucleic acid substitutions (C) and amino acid replacements (D) identified in

1523   *CsMLO4* (located on chromosome 2) from the same 32 hemp and drug-type cannabis cultivars as shown

1524   on panels A and B. On panels A and C (gDNA), pie charts display, for each SNP along the coding sequence,

1525   the genotype frequencies calculated among the 32 cultivars. Colored horizontal rectangles on panels A and

1526   C represent the exons, numbered from 1 to 15: dotted rectangles represent extracellular domains of the

1527   resulting protein, while blue and striped rectangles represent membrane and cytoplasmic domains,

1528   respectively. Genotype frequency color code on panels A and C: samples called as homozygous for the

1529   reference allele are depicted in orange, samples called as heterozygous are depicted in light blue, and

1530   samples called as homozygous for the alternate allele are depicted in green. On panels B and D (amino

1531   acid sequence), the gray vertical rectangles depict transmembrane domains and the solid black curves depict

1532   cytoplasmic and extracellular loops. Resulting amino acid replacements on panels B and D are color-coded

1533   according to the gain or loss of polarity: red circles with white cross display a replacement with an acidic

1534   amino acid, yellow circles with black hyphen represent a replacement with an alkaline amino acid, gray

1535   circles display a replacement with a non polar amino acid, and gray-striped white circles represent a
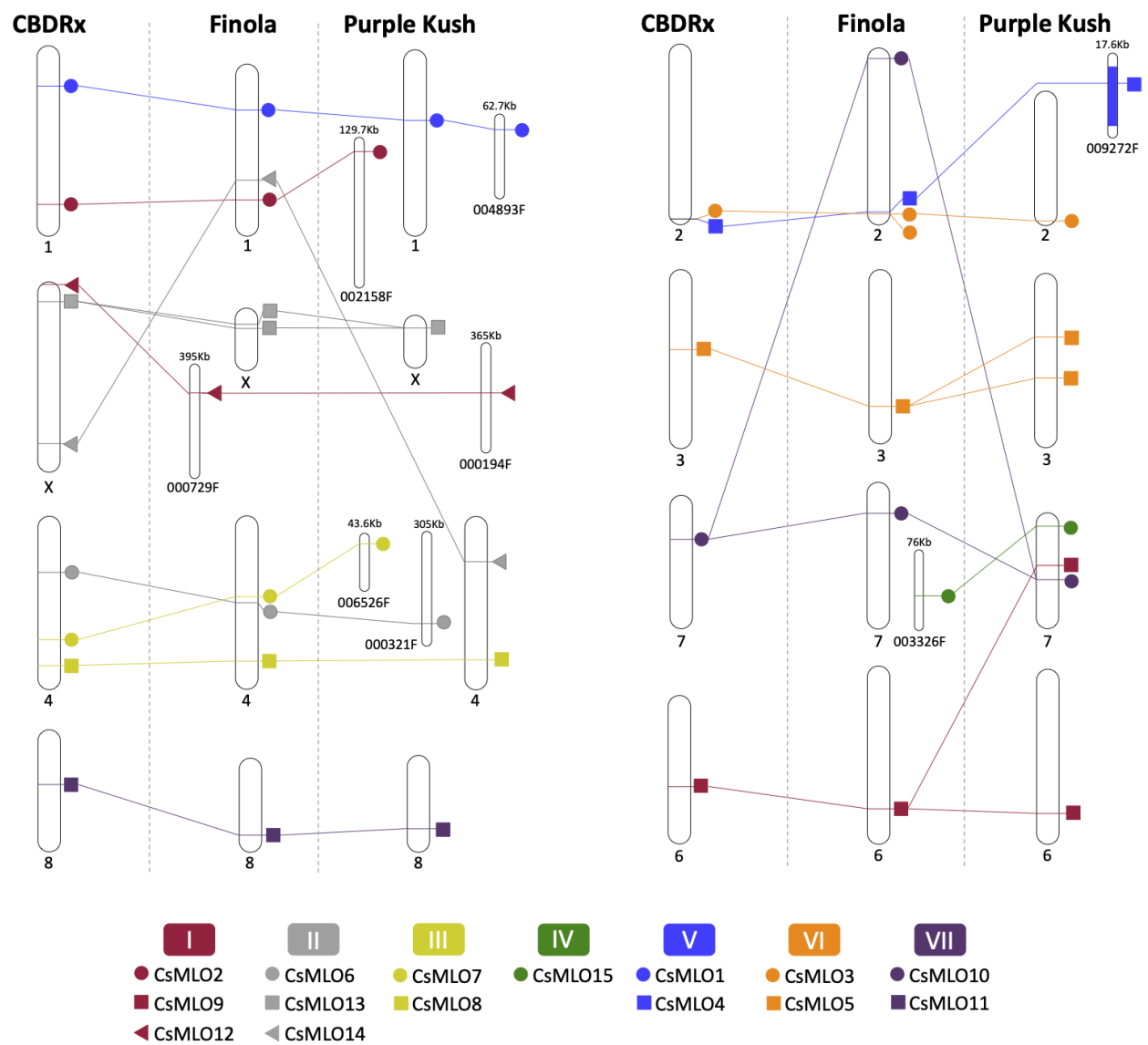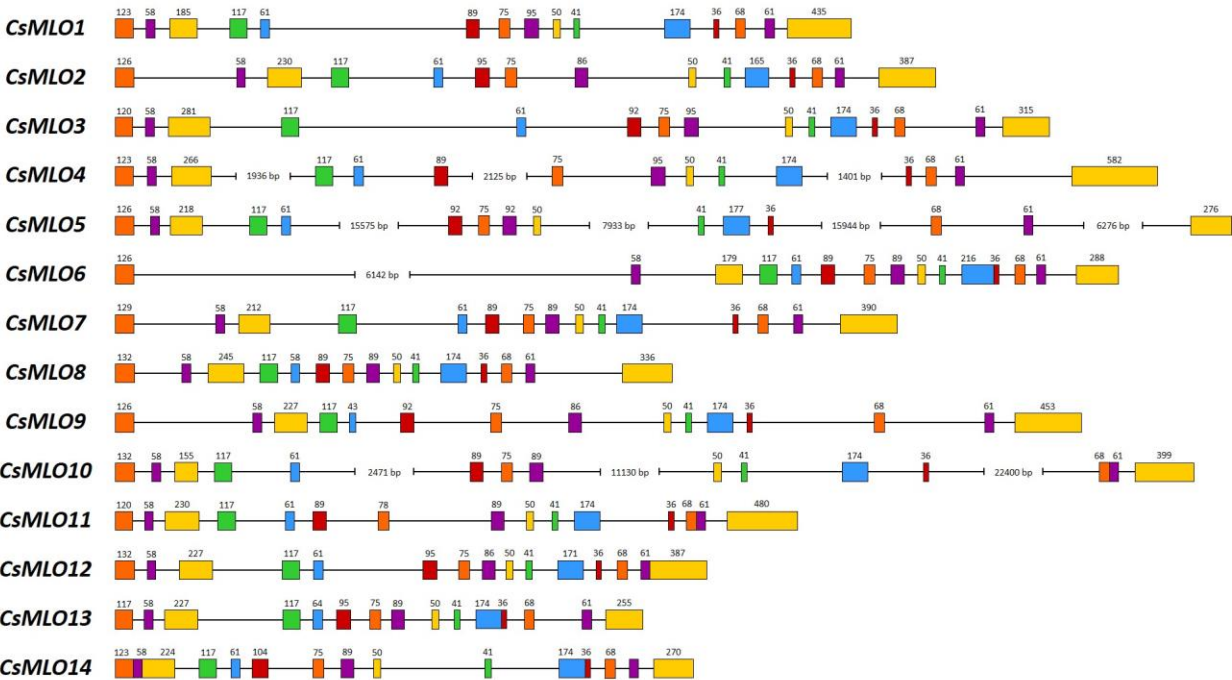
1536   replacement with a polar amino-acid.

1537

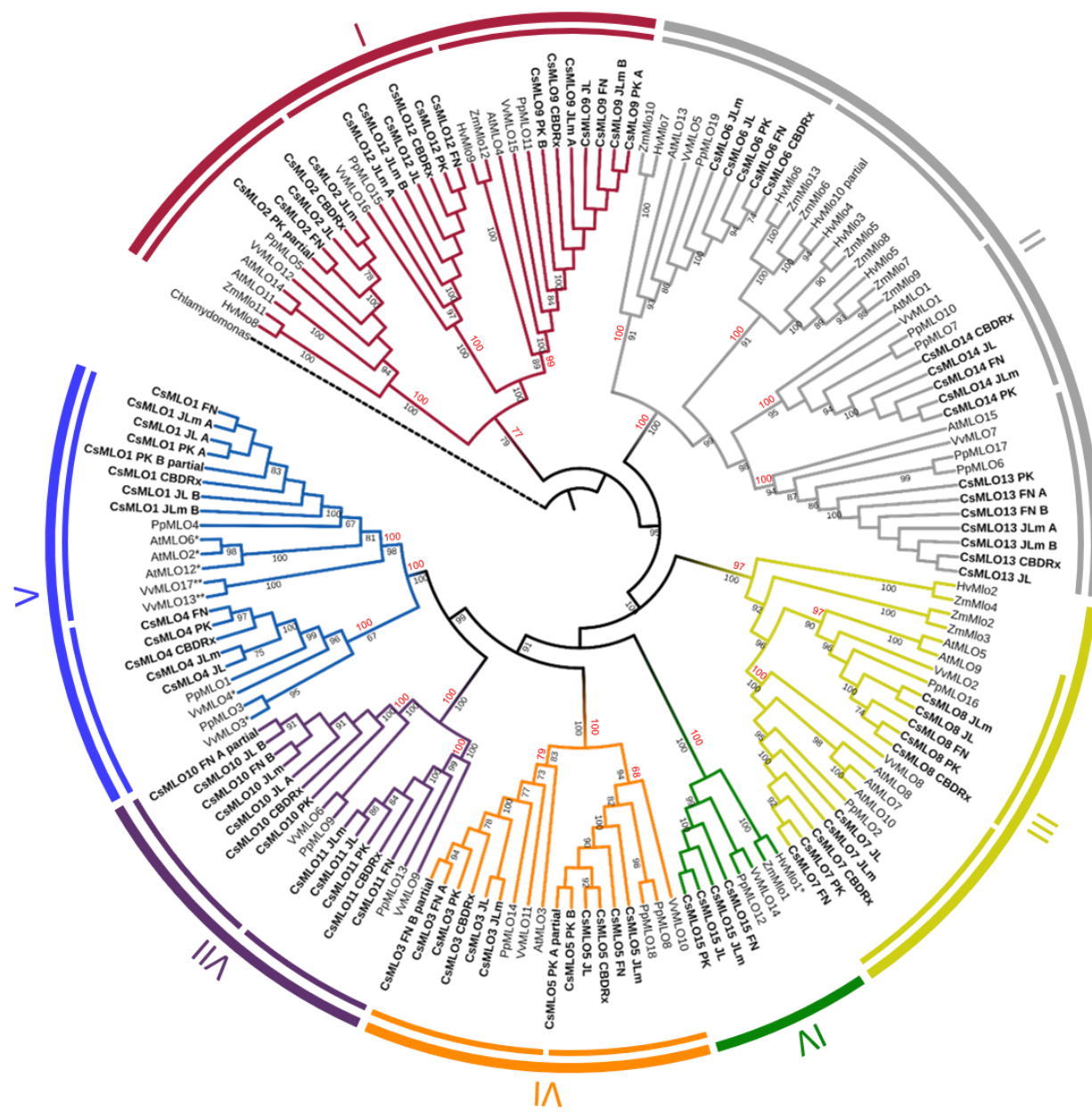1538   **Figure 1.**

1539

1540

Figure 2.

1541
1542 **Figure 3.**

1543
1544

Figure 4.
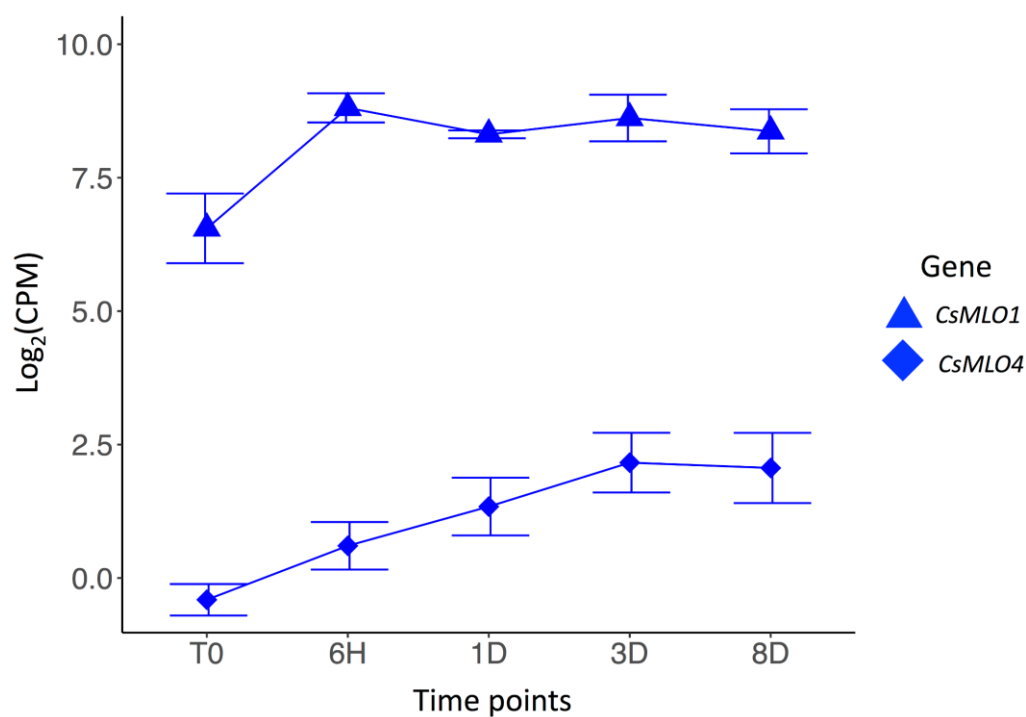
1545

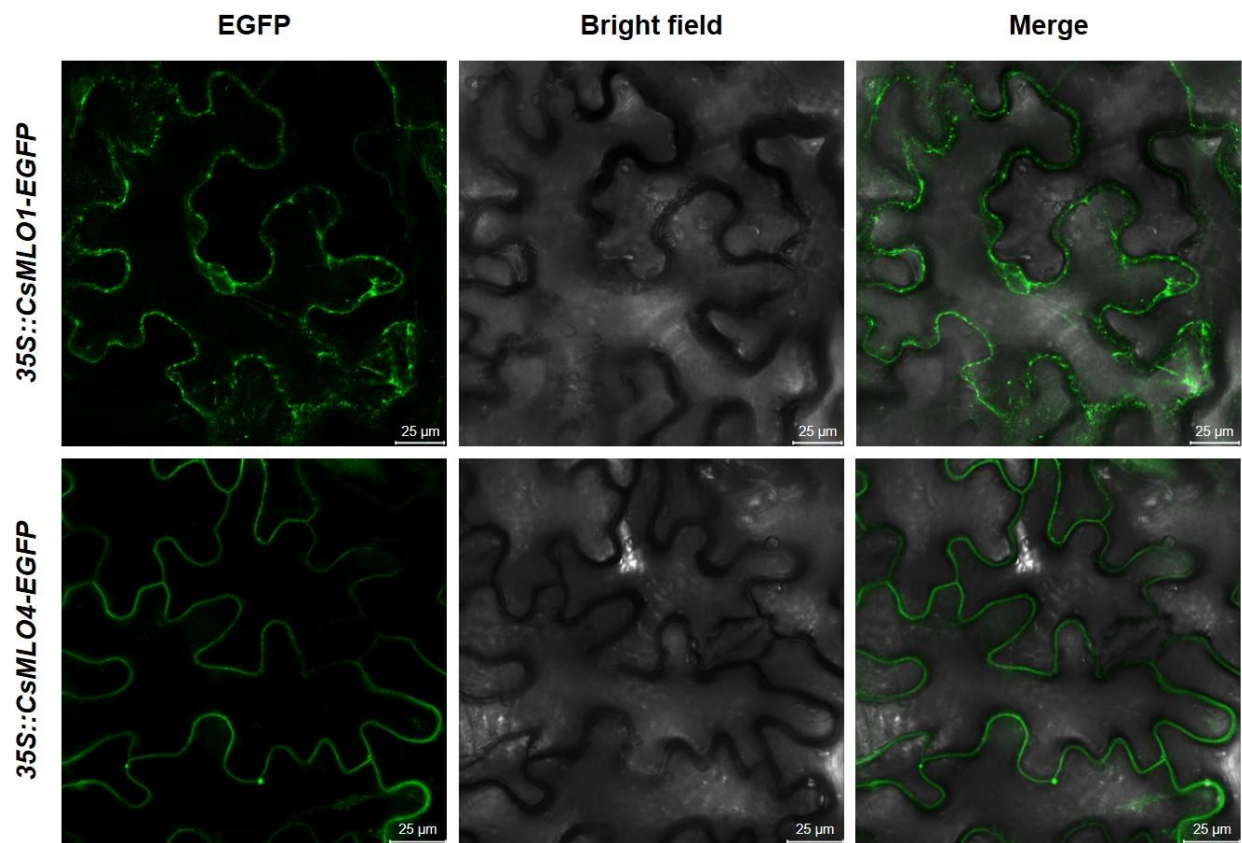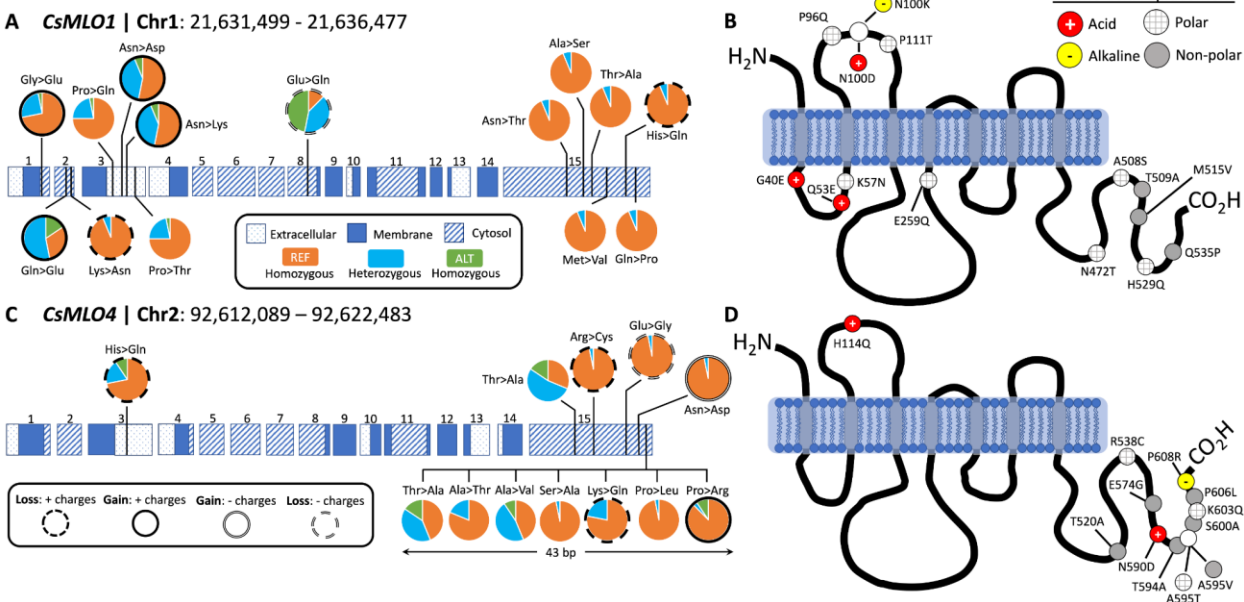1546   **Figure 5.**

Figure 6.

1549 **Supplementary materials**

1550 **Supplementary Table S1**: Members of the *CsMLO* gene family as predicted and manually curated

1551 in the CBDRx genome.

1552 **Supplementary Table S2**: Members of the *CsMLO* gene family as predicted and manually curated

1553 in the Jamaican Lion (female) genome.

1554 **Supplementary Table S3**: Members of the *CsMLO* gene family as predicted and manually curated

1555 in the Jamaican Lion (male) genome.

1556 **Supplementary Table S4**: Members of the *CsMLO* gene family as predicted and manually curated

1557 in the Purple Kush genome.

1558 **Supplementary Table S5**: Members of the *CsMLO* gene family as predicted and manually curated
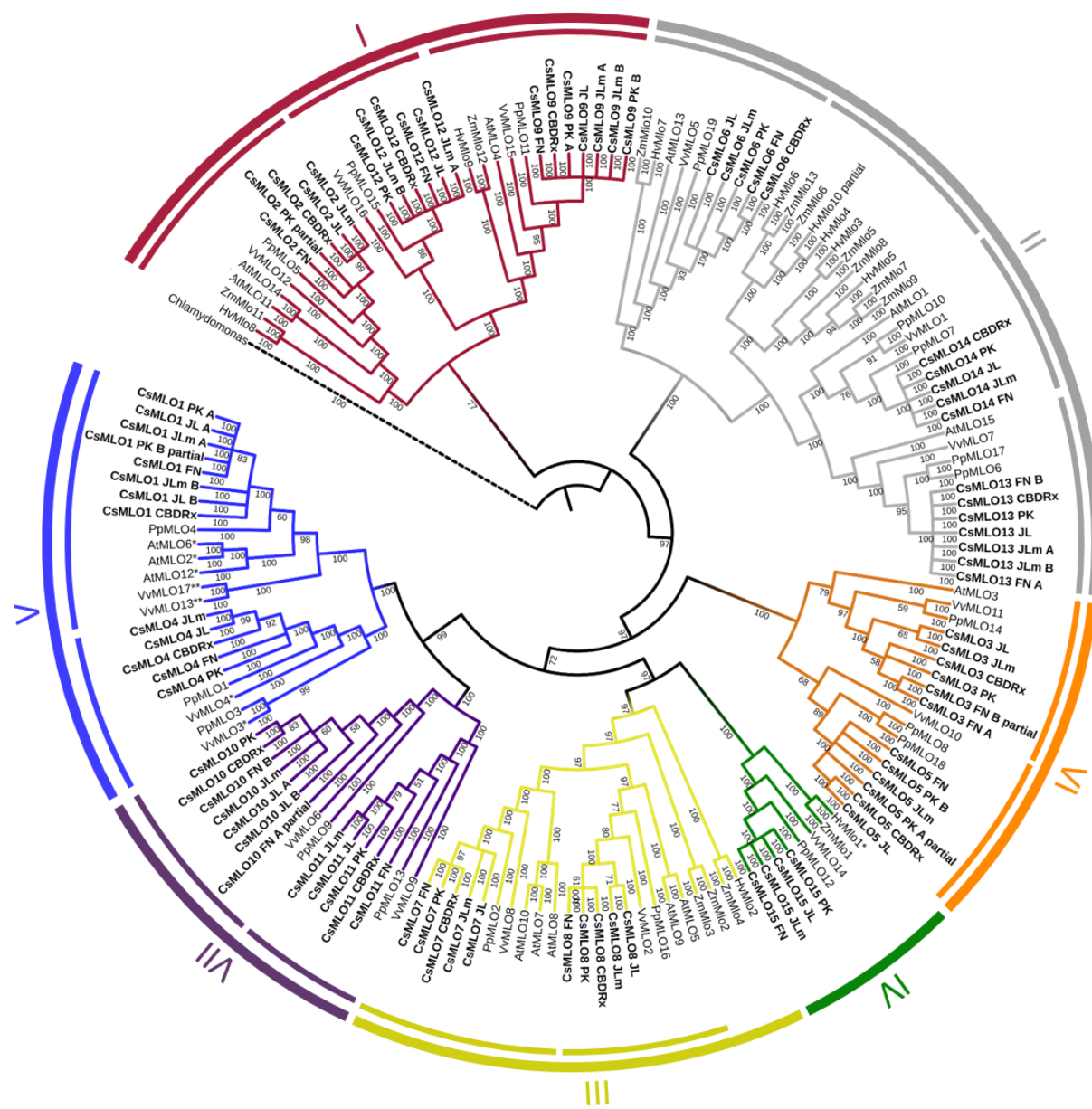
1559 in the Finola genome.

1560 **Supplementary Table S6**: Plant *cis*-acting regulatory elements identified in *CsMLOs* across five

1561 distinct cultivars.

1562 **Supplementary Table S7**: Non-synonymous SNPs identified in *CsMLO1* and *CsMLO4* of 32

1563 cannabis cultivars with their respective amino acid change in the resulting protein.
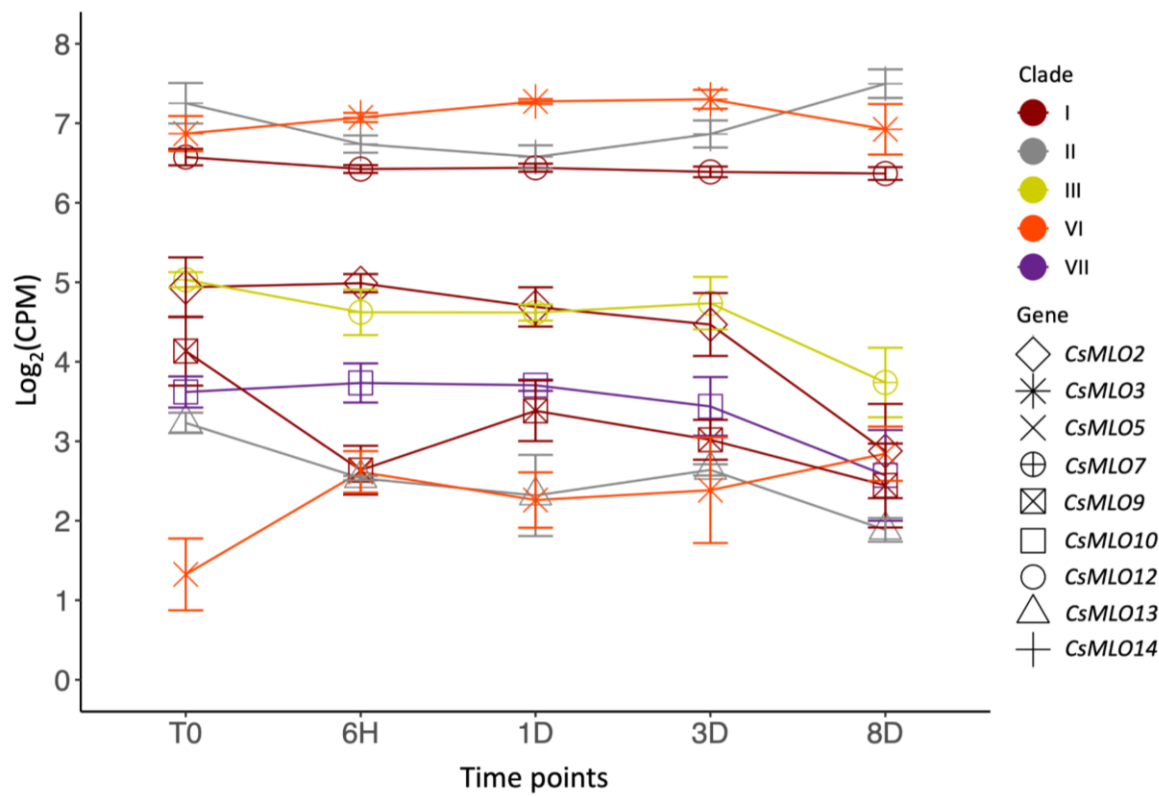
1564 **Supplementary Figure S1**: **Phylogenetic relationships of CsMLOs based on Bayesian**

1565 **inference analysis.** Phylogenetic tree of manually curated CsMLO proteins (bold) with MLO

1566 proteins from selected species (*Arabidopsis thaliana*, *Prunus persica*, *Vitis vinifera*, *Hordeum*

1567 *vulgare* and *Zea mays*). *Chlamydomonas reinhardtii* was used as an outgroup. Phylogenetic

1568 relationships were estimated using the MrBayes tool implemented on NGPhylogeny.fr, using

1569 default parameters. The seven defined clades are indicated, as well as potential subclades identified

1570 in this study (inner circles). Number on a node indicates the posterior probabilities of major clades

1571 and subclades. MLOs with one asterisk (*) have been experimentally demonstrated to be required

1572 for PM susceptibility (Büschges et al., 1997; Feechan et al., 2008; Wan et al., 2020), while MLOs

1573 with two asterisks (**) have been identified as main probable candidates for PM susceptibility

1574 (Pessina et al., 2016).

1575 **Supplementary Figure S2. Transcriptomic response of *CsMLO* genes from all clades, except**

1576 **from clade V, following laboratory-controlled inoculations with powdery mildew**. Specific

1577 *CsMLO* genes are represented by different symbols and the clade to which each gene belongs is

1578 displayed by different colors. Gene expression is displayed on the y-axis as the average logarithmic

1579 value of the Counts Per Million (logCPM) at each time point (displayed on the x-axis, n = 3 per

1580   time point). Time points: no infection / control (T0), 6 hours post-inoculation (6H), 24 hours post-

1581   inoculation (1D), 3 days post-inoculation (3D), and 8 days post-inoculation (8D).

1582   **Supplementary File S1**: Nucleotide alignment of the genomic sequences from all 86 *CsMLOs*.

1583   Exons and introns are separated by multiple gaps in the alignment, as is the case for the insertion

1584   in the first exon of *CsMLO5-JLm*.

1585   **Supplementary File S2**: Nucleotide alignment of the coding sequences from all 86 *CsMLOs*.

1586   **Supplementary File S3**: Amino acid alignment of the protein sequences from all 86 *CsMLOs*.

1587

1588 **Supplementary Figure S1.**

1589

1590 **Supplementary Figure S2.**