Genome-wide identification and characterization of major latex-like protein genes responsible for crop contamination in Cucurbita pepo

Fujita, Kentaro ; Chitose, Natsumi ; Chujo, Maho ; Komura, Shoya ; Sonoda, Chihiro ; Yoshida, Minami ; Inui, Hideyuki

(Citation)
Molecular Biology Reports,49(8):7773-7782

(Issue Date)
2022-08

(Resource Type)
journal article

(Version)
Accepted Manuscript

Genome-wide identification and characterization of *major latex-like protein* genes responsible for crop contamination in *Cucurbita pepo*

Kentaro Fujita[1] (kentaro.fujita1221@gmail.com) (https://orcid.org/0000-0002-8757-0094)

Natsumi Chitose[2]

Maho Chujo[2]

Shoya Komura[1] (shoya.komura@gmail.com)

Chihiro Sonoda[1]

Minami Yoshida[1]

Hideyuki Inui[1, 3, *] (hinui@kobe-u.ac.jp) (https://orcid.org/0000-0002-0981-2641)

[1]Graduate School of Agricultural Science, Kobe University, 1-1 Rokkodaicho, Nada-ku, Kobe, Hyogo 657-8501, Japan

[2]Faculty of Agriculture, Kobe University, 1-1 Rokkodaicho, Nada-ku, Kobe, Hyogo 657-8501, Japan

[3]Biosignal Research Center, Kobe University, 1-1 Rokkodaicho, Nada-ku, Kobe, Hyogo 657-8501, Japan

[*]Corresponding Author

Hideyuki Inui, Biosignal Research Center, Kobe University,

1-1 Rokkodaicho, Nada-ku, Kobe, Hyogo 657-8501, Japan

E-mail: hinui@kobe-u.ac.jp, Phone: +81-78-803-5863

Abstract

Background

Zucchini plants (*Cucurbita pepo*) accumulate persistent organic pollutants (POPs) at high concentrations in their aerial parts, and major latex-like proteins (MLPs) play crucial roles in their accumulation. MLPs bind to POPs in root cells, MLP-POP complexes are then translocated into xylem vessels, and POPs are transported to the aerial parts. We previously identified three CpMLP genes (*MLP-PG1*, *MLP-GR1*, and *MLP-GR3*) as transporting factors for POPs; however, other studies have shown that the genomes of several plant species contain more than 10 MLP genes, thus, further MLP genes responsible for POP accumulation may have been overlooked.

Methods and Results

Here, we investigated the number of CpMLP genes by performing a hidden Markov model search against the *C. pepo* genome database and characterized their effects on POP accumulation by performing the expression analysis in the organs and *in silico* structural analysis. The *C. pepo* genome contained 21 CpMLP genes, and several CpMLP genes, including *MLP-PG1* and *MLP-GR3*, were highly expressed in roots. 3D structural prediction showed that all examined CpMLPs contained a cavity with a hydrophobic region, which facilitated binding to POPs.

Conclusions

The present study provides insights regarding CpMLP genes responsible for POP accumulation.

Keywords

crop contamination; *Cucurbita pepo*; major latex-like protein; persistent organic pollutants

Abbreviations

ABA, abscisic acid; HMM, Hidden Markov Model; JA, jasmonic acid; MLP, major latex-like protein; POPs, persistent organic pollutants; SA, salicylic acid.

Introduction

The family Cucurbitaceae belongs to the order Cucurbitales in the rosid group, and its members are annual herbaceous vines and perennial lianas [1]. The genus *Cucurbita* comprises three subspecies, i.e., ssp. *fraterna*, ssp. *ovifera*, and ssp. *pepo* [2]. *C. pepo* ssp. *pepo* shows high accumulation of persistent organic pollutants (POPs), whereas *C. pepo* ssp. *ovifera* show low accumulation of POPs [3]. POPs are globally distributed and contaminate the environment owing to their persistence [4]. POP insecticides (dieldrin and hexachlorocyclohexane) currently occur in agricultural fields of many countries because large amounts of POP insecticides have been applied until their prohibition [5, 6]. POPs are highly toxic to humans, with effects such as neurotoxicity [7]; thus, elucidating the mechanisms underlying the accumulation of POPs should be addressed for safer crop production.

Major latex-like proteins (MLPs) bind to POPs absorbed into root cells, and MLP-POP complexes are translocated into xylem vessels, after which POPs are transported to the upper plant parts [8]. To date, MLP genes have been observed mainly in dicots such as *Arabidopsis thaliana* [9], *C. pepo* [10], cotton (*Gossypium hirsutum*) [11], soybean (*Glycine max*) [12], peach (*Prunus persica*) [13], and grapevine (*Vitis vinifera*) [14]. MLPs play a key role in leaf differentiation [15], stress responses to pathogens and drought [9, 11, 16], and biosynthesis of secondary metabolites [17].

MLPs are members of the Bet v 1 family, including pathogenesis-related protein class 10 (PR-10) [18]. Amino acid identity among proteins of this family is low; however, they contain three α-helices and seven β-sheets as common secondary structures, and they show similar 3D structures. The most prominent characteristic of proteins in the Bet v 1 family is an internal hydrophobic cavity formed by β1-β7 wrapped around the long α3 [18] through which MLPs can bind to various hydrophobic compounds such as steroids [19] and POPs [8]. Furthermore, MLPs have been detected in xylem and in phloem sap [20–23], indicating that CpMLPs transport POPs by binding them through the hydrophobic cavity. However, the number of MLP genes in the *C. pepo* genome remains unknown.

We previously identified *MLP-PG1*, *MLP-GR1*, and *MLP-GR3* [20], and MLP-PG1 and MLP-GR3 showed binding affinity for POPs [8]. A previous genome-wide analysis identified more than ten MLP genes in the genomes of *V. vinifera* [14], *Brassica rapa* [24], and *Malus domestica* [25], suggesting that several CpMLPs are involved in the accumulation of POPs. Therefore, MLPs responsible for POP transport in *C. pepo* may have been overlooked previously.

In this study, we investigated the number of CpMLP genes and identified CpMLP genes in the *C. pepo* genome. We characterized phylogenetic relationships, subcellular localization, conserved domains, conserved motifs, gene structures, and *cis*-acting regulatory elements. We further evaluated the expression levels of CpMLP genes in the organs of *C. pepo* ssp. *ovifera* and *C. pepo*. Moreover, we predicted 3D structures and calculated cavity size to test whether the cavities were sufficiently large to bind POPs. Our study provides insights regarding further CpMLPs associated with crop contamination in terms of subcellular localization, gene expression, and protein structural aspects.

Materials & Methods

Identification of CpMLP *genes in the C. pepo genome*

A Hidden Markov Model (HMM) of Bet v 1 (PF00407) was downloaded from Pfam (http://www.pfam.xfam.org/), and HMM Search (version 3.3.2) was performed against the *C. pepo* genome database (http://cucurbitgenomics.org/) with an E-value cutoff of 10^{-5}. A BlastP search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was also performed as a query of amino acid sequences of 25 AtMLPs identified in a previous study [14] against the *C. pepo* genome database with an E-value cut-off of 10^{-5}. Domain analysis was performed using SMART (http://smart.embl-heidelberg.de/) and INTERPROSCAN (http://www.ebi.ac.uk/interpro/) to confirm that candidate proteins contained the Bet v 1 domain, and those without the Bet v 1 domain were omitted.

Prediction of the subcellular localization

The subcellular localization of each CpMLP was determined using WoLF PSORT (https://wolfpsort.hgc.jp/).

Gene and protein analysis

A phylogenetic tree based on the amino acid sequences of 25 AtMLPs and 21 CpMLPs was constructed using the neighbor-joining method in MEGA 7.0 (http://www.megasoftware.net/), and 50 bootstrap replications were applied. Amino acid sequences of 21 CpMLPs were analyzed using NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to identify the location of Bet v 1 domain. Genes with introns and coding sequences were extracted from the *C. pepo* genome database and were visualized using the Gene Structure Display Server (version 2.0) (http://gsds.gao-lab.org/). The motifs in amino acid sequences of 21 CpMLPs were analyzed using MEME (version 5.4.1) (https://meme-suite.org/meme/tools/meme) with a maximum motif number of 10 and an optimum motif width of 6–50 residues. The 1-kbp regions upstream from the translation starting position in 21 CpMLP genes were extracted, and *cis*-acting regulatory elements related to plant hormones and root-specific expression were identified using the PLACE database [26].

RNA-sequencing (RNA-seq) analysis

Seeds of *C. pepo* ssp. *ovifera* cultivars ('Patty Green' [PG] and 'Starship' [ST]) and ssp. *pepo* cv. ('Gold Rush' [GR] and 'Magda' [MG]) were purchased from Johnny's Selected Seeds (Albion, ME, USA). The seed coat was peeled off, and seeds were incubated in tap water overnight at 4 °C and were sown in sterilized soil (Hyponex Japan Corp., Ltd., Osaka, Japan). The seedlings were cultivated under a 16/8 h light/dark cycle at 26 °C for 40 days. The roots were collected and were ground in liquid nitrogen, and total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). A

sequencing library was constructed according to the cDNA Rapid Library Preparation Method Manual, and RNA sequencing (GS-FLX+) (Roche, Basel, Switzerland) was performed by Takara Bio Inc. (Shiga, Japan).

Expression analysis

The seed coat of *C. pepo* ssp. *ovifera* cv. PG and ssp. *pepo* cv. GR was peeled off. The seedlings were cultivated under a 16/8 h light/dark cycle at 26 °C as described above. After 13 days, the organs (leaves and roots) were collected and were ground in liquid nitrogen. Total RNA extraction and quantitative reverse-transcription PCR (qRT-PCR) were performed as previously described [27].

Construction of 3D structures

The predicted structures of the CpMLPs were constructed using AlphaFold2 [28]. Cavity volume was calculated using CASTp (http://sts.bioe.uic.edu/castp/calculation.html).

Results

Identification of CpMLP *genes in C. pepo genome*

An HMM search was performed using the HMM profile of Bet v 1 (PF00407) as a query against the *C. pepo* genome database, and 26 candidate CpMLPs were detected. A BlastP search was performed using 25 AtMLPs as a query against the *C. pepo* genome database [14], and 22 candidate CpMLPs were detected. After removing overlapping proteins in the HMM and BlastP searches, 30 candidate CpMLPs were identified. The presence of the Bet v 1 domain in the amino acid sequences was determined using SMART and INTERPRO, and 21 putative CpMLPs were identified (Table 1, Suppl. table S1).

145 The amino acids and molecular weights of the 21 CpMLPs were assessed based on their DNA
146 sequences (Table 1). The amino acids and molecular weights of the 15 CpMLPs were relatively similar,
147 ranging from 142 to 205 amino acid sequences and from 15,392 to 23,552 Da, respectively. CpMLP6
148 and CpMLP10 showed few amino acids and low molecular weights. In contrast, CpMLP5, CpMLP11,
149 CpMLP13, and CpMLP19 contained many amino acids and showed markedly higher molecular
150 weights. CpMLP genes were located on chromosomes 0, 1, 2, 3, 12, 14, 16, and 17. Chromosome 1
151 harbored the most CpMLP genes (eight CpMLP genes), and seven of them were included in subfamily
152 VI (Fig. 1a), however, their amino acid content and molecular weights varied widely. CpMLP genes
153 located on chromosomes 2 and 3 were included in subfamily IV, and CpMLP genes located on
154 chromosomes 14, 16, and 17 were included in subfamily I (Fig. 1a).
155
156 *Phylogenetic tree of CpMLPs*
157 To classify the 21 CpMLPs with 25 AtMLPs, a phylogenetic tree was constructed using the neighbor-
158 joining method (Fig. 1a). AtMLPs and CpMLPs were divided into seven subfamilies: I (8 CpMLPs),
159 II (AtMLP and CpMLP), III (10 AtMLPs), IV (5 CpMLPs), V (2 AtMLPs), VI (7 CpMLP), and VII
160 (12 AtMLPs). CpMLPs clustered in subfamilies I, IV, and VI, except for CpMLP3, which was grouped
161 in subfamily II. The AtMLPs and CpMLPs were clustered, except for subfamily II.
162
163 *Subcellular localization*
164 The subcellular localization of CpMLPs was predicted using WoLF PSORT, and all CpMLPs, except
165 for CpMLP5, CpMLP10, and CpMLP14, showed the highest scores in the cytosol (Fig. 1b); CpMLP5,
166 CpMLP10, and CpMLP14 produced the highest scores in the plasma membrane, mitochondria, and
167 nucleus, respectively. CpMLP5 did not produce a score in the cytosol. Most CpMLPs produced a score
168 in the chloroplast, extracellular space, and nucleus, apart from the cytosol.

169

170 *Gene structures and conserved motifs of CpMLPs*

171 The exon-intron structures of CpMLP genes were determined, and all CpMLP genes contained at least
172 two exons and one intron (Fig. 2a). Fifteen of the 21 CpMLP genes contained two exons divided by
173 one exon. The CpMLP5, CpMLP11, and CpMLP13 contained at least seven exons and six introns and
174 differed from other CpMLP genes. Although CpMLP5, CpMLP11, and CpMLP13 were distributed in
175 subfamilies VI (CpMLP5) and IV (CpMLP11 and CpMLP13), these unique gene structures did not
176 depend on the subfamily.

177

178 *Domain and motif structures in CpMLPs*

179 Eighteen of the 21 CpMLPs contained one Bet v 1 domain in their amino acid sequence (Fig. 2b).
180 CpMLP11 and CpMLP19 contain three and two Bet v 1 domains, respectively. CpMLP13 contained
181 seven Bet v 1 domains. MEME, a motif discovery tool, was used to investigate conserved motifs in
182 the amino acid sequences of the CpMLPs. Motif 1
183 (GDLFEHFKVFKVVYKVVEKGPNSCJVVLTIEYEKLEEGAAN) was the most frequent in
184 CpMLPs, and 13 CpMLPs contained this motif in their posterior sequences (Fig. 2c). Motifs 7
185 (FKERVEFDDEKFTIVLVGLE) and 9 (PYKYJDLMNKJTKDI) were the second most frequent, and
186 12 CpMLPs contained these in their posterior sequences. CpMLP4, CpMLP5, CpMLP7, CpMLP8,
187 CpMLP9, CpMLP10, CpMLP14, and CpMLP15 contained these three motifs and were included in
188 subfamilies IV (CpMLP4 and CpMLP5) and VI (CpMLP7, CpMLP8, CpMLP9, CpMLP10,
189 CpMLP14, and CpMLP15) (Fig. 1a).

190

191 *Cis-acting regulatory elements in the promoter region of* CpMLP *genes*

192 The 1-kbp region upstream of the translation initiation site of 21 CpMLP genes was analyzed using
193 PLACE to identify potential *cis*-acting regulatory elements related to plant hormones and root-specific
194 expression in the promoter region. PLACE identified seven *cis*-acting regulatory elements of plant
195 hormones: abscic acid (ABA)-responsive, auxin-responsive, cytokinin-responsive, ethylene-
196 responsive, gibberellin-responsive, JA-responsive, and salicylic acid (SA)-responsive elements (Fig.
197 3a). All CpMLP genes, except for *CpMLP4*, contained ABA-responsive elements. All CpMLP genes
198 contained gibberellin-responsive and SA-responsive elements. In contrast, the number of *cis*-acting
199 regulatory elements related to other plant hormones (auxin-responsive, cytokinin-responsive,
200 ethylene-responsive, and JA-responsive elements) was relatively low. All CpMLP genes contained
201 root-specific expression elements.
202
203 *Expression level of* CpMLP *genes in C. pepo subspecies*
204 RNA-seq using roots of *C. pepo* ssp. *ovifera* and ssp. *pepo* cultivars was performed to identify CpMLP
205 genes with high expression. MLP genes that showed the most homologous sequences with the 21
206 CpMLP genes were selected. *CpMLP6* and *CpMLP12* showed high read counts in *C. pepo* ssp. *ovifera*
207 cultivars and a pattern similar to that of *MLP-PG1* (Fig. 3b). In contrast, *CpMLP8* showed high read
208 counts in *C. pepo* ssp. *pepo* cultivars and a pattern similar to that of *MLP-GR3* (Fig. 3b). Sixteen of
209 the 21 CpMLP genes produced < 10 reads in *C. pepo* ssp. *ovifera* and ssp. *pepo*. The examined
210 cultivars did not contain sequences matching *CpMLP3* and *CpMLP11*.
211 As *CpMLP4*, *CpMLP6*, *CpMLP8*, *CpMLP9*, *CpMLP12*, and *CpMLP21* showed > 10 reads in the roots
212 of at least one cultivar, the expression of these CpMLP genes and of *MLP-PG1* and *MLP-GR3* were
213 investigated. *MLP-PG1* and *MLP-GR3* showed remarkably high expression in roots of *C. pepo* ssp.
214 *pepo* cv. GR (Fig. 3c). *CpMLP8* and *CpMLP9* had higher expression in roots of *C. pepo* ssp. *pepo* cv.
215 GR, compared with those of *C. pepo* ssp. *ovifera* cv. PG, which corresponded with the RNA-seq results.

However, *CpMLP6* showed higher expression in roots of *C. pepo* ssp. *pepo* cv. GR compared with *C. pepo* ssp. *ovifera* cv. PG, in contrast to the RNA-seq results. *CpMLP4*, *CpMLP12*, and *CpMLP21* were not highly expressed in roots. All examined MLP genes showed remarkably lower expression in leaves than in roots.

3D structures of CpMLPs

The predicted 3D structures of the CpMLPs were constructed using AlphaFold2 (Suppl. fig 1). CpMLPs of approximately 17 kDa showed canonical MLP secondary structures: three α-helices and seven β-sheets. CpMLPs in the same subfamily did not consistently show the same structure (Fig. 1a). All CpMLPs contained a hydrophobic region in their cavities (Suppl. fig 1). CpMLPs with molecular weights twice as high as 17 kDa tended to have a large cavity volume (Table 2). Several CpMLPs with similar molecular weights showed similar cavity volumes. CpMLP7, CpMLP8, and CpMLP9 in subfamily VI showed similar cavity volumes. However, several CpMLPs showed different cavity volumes, although their molecular weights and structures were similar; the cavity volume of CpMLP1 was four-fold smaller than that of CpMLP2. Because CpMLP13 had an uncharacterized structure, its structure was not assessed using AlphaFold2.

Discussion

The objective of the present study was to identify CpMLP genes in the *C. pepo* genome and to produce insights into their effects on the accumulation of POPs. *C. pepo* ssp. *pepo* accumulates POPs in its aerial parts at high concentrations [3]. We previously identified three CpMLPs as crucial factors for their accumulation [20]. Previous studies have shown that dicot plant genomes contain more than ten MLP genes [14]; however, the number of CpMLPs that can accumulate POPs remains elusive.

239 Therefore, we investigated the number of CpMLP genes in the *C. pepo* genome database and predicted
240 their effects on POP accumulation.
241 HMM and BlastP searches showed that the *C. pepo* genome contained 21 CpMLP genes with single
242 or several Bet v 1 domains (Table 1, Fig. 1b). Previous studies have shown that the genomes of *B.*
243 *rapa*, *M. domestica*, and *V. vinifera* contain 31, 36, and 14 MLP genes, respectively [14, 24, 25]. The
244 different number of MLP genes in plant species is presumably associated with genome size (*B. rapa*:
245 442.9 M, *M. domestica*: 687 M, and *V. vinifera*: 487 M) and the respective biological functions [14,
246 25, 29]. AtMLPs and CpMLPs were divided into seven subfamilies based on their amino acid
247 sequences, and all subfamilies, except for subfamily II, were exclusive to either AtMLPs or CpMLPs
248 (Fig. 1a). This suggested that MLP genes diversified after the speciation of *A. thaliana* and *C. pepo*.
249 We predicted the subcellular localization of CpMLPs (Fig. 1b). Interestingly, most CpMLPs had the
250 highest score in the cytosol. POPs absorbed into the roots are diffused in the plasma membrane and
251 are transported to the endodermis and pericycle without being blocked by the Casparian strip, a
252 diffusion barrier in the endodermis attaching to the intercellular walls [30, 31]. Therefore, CpMLPs
253 localized in the cytosol bind to and solubilized POPs in the cytosol. As a result, their complexes
254 reached xylem vessels and were transported to the aerial parts. Furthermore, CpMLPs localized in the
255 extracellular region tended to produce similar scores in the endoplasmic reticulum and Golgi apparatus
256 (Fig. 1b). CpMLP genes are highly expressed in the roots, and CpMLPs are detected in the xylem sap
257 and roots [8]. This clearly shows that CpMLPs are secreted from root cells to the outside of the cells.
258 The endoplasmic reticulum and Golgi apparatus play crucial roles in protein secretion in plants [32].
259 This suggests that CpMLPs are secreted from the endoplasmic reticulum and Golgi apparatus into the
260 extracellular space. However, we did not show the the subcellular localization of CpMLPs *in vivo*, and
261 future work will investigate their subcellular localization by the injection of *MLP::GFP* constructs
262 using the laser confocal microscopy.

263 All CpMLP genes contained *cis*-acting regulatory elements associated with plant hormone response
264 in the promoter region, indicating that most CpMLP genes were regulated by plant hormones such as
265 ABA, gibberellin, and SA (Fig. 3a). Previous studies have shown that MLPs confer drought tolerance
266 through mediating the ABA signaling pathway. ABA downregulates *MLP43* as negative feedback, and
267 MLP43 functions as an ABA regulator through its interaction with SnRK2 and ABF1 [9]. ABA
268 downregulates *TaSTP*, an *MLP* identified in wheat (*Triticum aestivum*), and TaSTP is thought to confer
269 drought tolerance through its interaction with TaDIS1 [33]. Furthermore, ABA receptors show
270 structural similarities to Bet v 1 [34, 35]. Therefore, MLPs exhibit ABA responses in terms of gene
271 expression and receptors. JA upregulates MLP genes in several plants, including *C. pepo* [16],
272 mulberry (*Morus multicaulis*) [22], and tobacco (*Nicotiana benthamiana*) [36]. However, there are
273 only few reports on upregulation of MLP genes by SA. Because CpMLP genes contained several SA-
274 responsive elements, these results are unexpected, with respect to those of previous reports (Fig. 3a).
275 PR-10 genes tend to be upregulated by SA [37], and CpMLP genes show upregulation mechanisms
276 similar to those of PR-10 genes, rather those of typical MLP genes. MLP genes identified previously
277 in *C. pepo* [8, 16] were highly expressed in roots, thus, the observed low expression levels of most of
278 the 21 CpMLP genes were unexpected (Fig. 3b and c). However, the expression of each MLP gene
279 differs between organs in several plants [14, 25]. For example, *VvMLP5* and *VvMLP14* were highly
280 expressed in roots and leaves, respectively.
281 All examined CpMLPs showed an internal cavity with a hydrophobic region (Suppl. fig 1). The cavity
282 volume differed remarkably among CpMLPs. MLPs, whose structures were previously identified, also
283 revealed large differences in cavity volume. For example, the cavity volume of ginseng (*Panax*
284 *ginseng*) MLP is six-fold smaller than that of MLP28 of *A. thaliana* [38]. This causes a kink in the
285 long α3 toward the core β-sheets or the existence of a long loop [19, 39]. CpMLP17 and CpMLP18
286 contained a similar amino acid sequence but showed different cavity volumes (Table 2). The α3 in

CpMLP17 was more kinked than that in CpMLP18, leading to a difference in cavity volume (Suppl. fig 1). However, these events were not always observed in CpMLPs. CpMLP1 and CpMLP2, with similar amino acid sequences, did not show a kink of α3 and did not possess a long loop like MLP28, but showed remarkable cavity volume differences (Table 2). The loop between β2 and β3 of CpMLP2 stretches from the core β-sheets (Suppl. fig 1), indicating higher cavity volume in CpMLP2.

Only genes with a sequence similar to that of *CpMLP8* were highly expressed in the roots of *C. pepo* ssp. *pepo* cultivars, whereas others were not expressed at high levels (Fig. 3b). The expression pattern of *CpMLP8* showed a similar tendency to that of *MLP-GR3*, which is a crucial factor for the transport of POPs in *C. pepo* ssp. *pepo*. CpMLP8 was predicted to be mainly localized in the cytosol and secreted into the extracellular fluid (Fig. 1b). The cavity size of CpMLP8 was larger than that of MLP-PG1, with the ability to bind POPs (Table 2) [8], indicating that CpMLP8 has a sufficiently large cavity to bind to POPs. Taken together, *CpMLP8* is expressed in roots, CpMLP8 binds to POPs in the cytosol of root cells, and their complexes are secreted into the extracellular fluid. Therefore, CpMLP8 played a crucial role in POP accumulation. To further our research we are planning to perform the cloning of CpMLP8 for *in vitro* assay of the binding to POPs and identify CpMLP8 in xylem sap of *C. pepo* by the amino acid sequence.

Conclusion

In the present study, we identified 21 CpMLP genes in the *C. pepo* genome and characterized their effect on transport of POPs in the subcellular localization, expression level, and 3D structure. CpMLP8 is another crucial factor in POP transport. Therefore, upregulation or downregulation of *CpMLP8* leads to phytoremediation and safe crop production, respectively. However, we cannot say with certainty whether CpMLP8 transports POPs in *C. pepo*; thus, *in vitro* and *in vivo* studies on CpMLP8-mediated transport of POPs should be conducted in the future.

311

312 **Author contribution**

313 **Conceptualization**: Kentaro Fujita; **Methodology**: Kentaro Fujita, Chitose Natsumi. Maho Chujo,
314 Shoya Komura, Chihiro Sonoda, Minami Yoshida, and Hideyuki Inui; **Formal analysis and**
315 **investigation**: Kentaro Fujita, Chitose Natsumi. Maho Chujo, Shoya Komura, Chihiro Sonoda,
316 Minami Yoshida, and Hideyuki Inui; **Writing - original draft preparation**: Kentaro Fujita and
317 Hideyuki Inui; **Writing - review and editing**: Kentaro Fujita, Chitose Natsumi. Maho Chujo, Shoya
318 Komura, Chihiro Sonoda, Minami Yoshida, and Hideyuki Inui; **Funding acquisition**: Kentaro Fujita
319 and Hideyuki Inui; **Resources**: Kentaro Fujita and Hideyuki Inui; **Supervision**: Hideyuki Inui

320

321 **Fundings**

322 This study was supported in part by SPRING of Japan Science and Technology Agency to KF and a
323 Grant-in-Aid for Scientific Research A from the Ministry of Education, Culture, Sports, and
324 Technology of Japan (No. 23241028) to HI.

325

326 **Acknowledgements**

327 We are thankful to Dr. Kentaro Yoshida (Graduate School of Agricultural Science, Kyoto University)
328 and Dr. Yoshihiro Matsuoka (Graduate School of Agricultural Science, Kobe University) for the
329 fruitful discussions in which helped writing this manuscript.

330

331 **Ethical approval**

332 This article does not contain any studies with human participants performed by any of the authors.

333

334 **Conflict of interest**

335 The authors have no conflict of interests to declare.

336

337 **References**

338 1. Guo J, Xu W, Hu Y, et al (2020) Phylotranscriptomics in Cucurbitaceae Reveal Multiple Whole-
339 Genome Duplications and Key Morphological and Molecular Innovations. Mol Plant 13:1117–
340 1133. https://doi.org/10.1016/j.molp.2020.05.011

341 2. Montero-Pau J, Blanca J, Bombarely A, et al (2018) De novo assembly of the zucchini genome
342 reveals a whole-genome duplication associated with the origin of the Cucurbita genus. Plant
343 Biotechnol J 16:1161–1171. https://doi.org/10.1111/pbi.12860

344 3. Inui H, Wakai T, Gion K, et al (2008) Differential uptake for dioxin-like compounds by zucchini
345 subspecies. Chemosphere 73:1602–1607. https://doi.org/10.1016/j.chemosphere.2008.08.013

346 4. Bartrons M, Catalan J, Penuelas J (2016) Spatial And Temporal Trends Of Organic Pollutants In
347 Vegetation From Remote And Rural Areas. Sci Rep 6:25446. https://doi.org/10.1038/srep25446

348 5. Jorgenson JL (2001) Aldrin and Dieldrin: A Review of Research on Their Production,
349 Environmental Deposition and Fate, Bioaccumulation, Toxicology, and Epidemiology in the
350 United States. Environ Health Perspect 109:113. https://doi.org/10.2307/3434852

351 6. Vijgen J, Abhilash PC, Li YF, et al (2011) Hexachlorocyclohexane (HCH) as new Stockholm
352 Convention POPs-a global perspective on the management of Lindane and its waste isomers.
353 Environ Sci Pollut Res 18:152–162. https://doi.org/10.1007/s11356-010-0417-9

354 7. Guo W, Pan B, Sakkiah S, et al (2019) Persistent Organic Pollutants in Food: Contamination
355 Sources, Health Effects and Detection Methods. Int J Environ Res Public Health 16:4361.
356 https://doi.org/https://doi.org/10.3390/ijerph16224361

8. Goto J, Iwabuchi A, Yoshihara R, et al (2019) Uptake mechanisms of polychlorinated biphenyls in *Cucurbita pepo* via xylem sap containing major latex-like proteins. Environ Exp Bot 162:399–405. https://doi.org/10.1016/j.envexpbot.2019.03.019

9. Wang Y, Yang L, Chen X, et al (2016) *Major latex protein-like protein 43* (*MLP43*) functions as a positive regulator during abscisic acid responses and confers drought tolerance in *Arabidopsis thaliana*. J Exp Bot 67:421–434. https://doi.org/10.1093/jxb/erv477

10. Iwabuchi A, Katte N, Suwa M, et al (2020) Factors regulating the differential uptake of persistent organic pollutants in cucurbits and non-cucurbits. J Plant Physiol 245:153094. https://doi.org/10.1016/j.jplph.2019.153094

11. Yang CL, Liang S, Wang HY, et al (2015) Cotton major latex protein 28 functions as a positive regulator of the ethylene responsive factor 6 in defense against *Verticillium dahliae*. Mol Plant 8:399–411. https://doi.org/10.1016/j.molp.2014.11.023

12. Guo D, Wong WS, Xu WZ, et al (2011) *Cis-cinnamic acid-enhanced 1* gene plays a role in regulation of Arabidopsis bolting. Plant Mol Biol 75:481–495. https://doi.org/10.1007/s11103-011-9746-4

13. Ruperti B, Bonghi C, Ziliotto F, et al (2002) Characterization of a major latex protein (MLP) gene down-regulated by ethylene during peach fruitlet abscission. Plant Sci 163:265–272. https://doi.org/10.1016/S0168-9452(02)00094-8

14. Zhang N, Li R, Shen W, et al (2018) Genome-wide evolutionary characterization and expression analyses of major latex protein (*MLP*) family genes in *Vitis vinifera*. Mol Genet Genomics 293:1061–1075. https://doi.org/10.1007/s00438-018-1440-7

15. Litholdo CG, Parker BL, Eamens AL, et al (2016) Proteomic Identification of Putative MicroRNA394 Target Genes in *Arabidopsis thaliana* Identifies Major Latex Protein Family

Members Critical for Normal Development. Mol Cell Proteomics 15:2033–2047. https://doi.org/10.1074/mcp.M115.053124

16. Fujita K, Asuke S, Isono E, et al (2022) MLP-PG1, a major latex-like protein identified in *Cucurbita pepo*, confers resistance through the induction of pathogenesis-related genes. Planta 255:10. https://doi.org/10.1007/s00425-021-03795-x

17. Lichman BR, Godden GT, Hamilton JP, et al (2020) The evolutionary origins of the cat attractant nepetalactone in catnip. Sci Adv 6:20. https://doi.org/10.1126/sciadv.aba0721

18. Fernandes H, Michalska K, Sikorski M, Jaskolski M (2013) Structural and functional aspects of PR-10 proteins. FEBS J 280:1169–1199. https://doi.org/10.1111/febs.12114

19. Lytle BL, Song J, de la Cruz NB, et al (2009) Structures of two *Arabidopsis thaliana* major latex proteins represent novel helix-grip folds. Proteins Struct Funct Bioinforma 76:237–243. https://doi.org/10.1002/prot.22396

20. Inui H, Sawada M, Goto J, et al (2013) A Major Latex-Like Protein Is a Key Factor in Crop Contamination by Persistent Organic Pollutants. Plant Physiol 161:2128–2135. https://doi.org/10.1104/pp.112.213645

21. Carella P, Merl-Pham J, Wilson DC, et al (2016) Comparative Proteomics Analysis of Phloem Exudates Collected During the Induction of Systemic Acquired Resistance. Plant Physiol 171:1495–1510. https://doi.org/10.1104/pp.16.00269

22. Gai YP, Yuan SS, Liu ZY, et al (2018) Integrated phloem sap mRNA and protein expression analysis reveals phytoplasma-infection responses in mulberry. Mol Cell Proteomics 17:1702–1719. https://doi.org/10.1074/mcp.RA118.000670

23. Li P, Chen L, Zhou Y, et al (2013) Brassinosteroids-Induced Systemic Stress Tolerance was Associated with Increased Transcripts of Several Defence-Related Genes in the Phloem in *Cucumis sativus*. PLoS One 8:1–8. https://doi.org/10.1371/journal.pone.0066582

24. Zeng J, Ruan Y, Liu B, et al (2020) Genome-wide identification and abiotic stress-responsive expression of *MLP* family genes in *Brassica rapa*. Gene Reports 21:100919. https://doi.org/10.1016/j.genrep.2020.100919

25. Yuan G, He S, Bian S, et al (2020) Genome-wide identification and expression analysis of major latex protein (*MLP*) family genes in the apple (*Malus domestica* Borkh.) genome. Gene 733:144275. https://doi.org/10.1016/j.gene.2019.144275

26. Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res 27:297–300. https://doi.org/10.1093/nar/27.1.297

27. Inui H, Katte N, Goto J, Iwabuchi A (2020) High temperatures promote the uptake of hydrophobic pollutants by *Cucurbita pepo via* altered gene expression levels of major latex-like proteins. J Pestic Sci 45:75–80. https://doi.org/10.1584/jpestics.d19-065

28. Jumper J, Evans R, Pritzel A, et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

29. Zhang L, Cai X, Wu J, et al (2018) Improved Brassica rapa reference genome by single-molecule sequencing and chromosome conformation capture technologies. Hortic Res 5:. https://doi.org/10.1038/s41438-018-0071-9

30. Roppolo D, De Rybel B, Tendon VD, et al (2011) A novel protein family mediates Casparian strip formation in the endodermis. Nature 473:381–384. https://doi.org/10.1038/nature10070

31. Yamazaki K, Tsuruta H, Inui H (2015) Different uptake pathways between hydrophilic and hydrophobic compounds in lateral roots of *Cucurbita pepo*. J Pestic Sci 40:99–105. https://doi.org/10.1584/jpestics.D14-081

32. Drakakaki G, Dandekar A (2013) Protein secretion: How many secretory routes does a plant cell have? Plant Sci 203–204:74–78. https://doi.org/10.1016/j.plantsci.2012.12.017

33. Lv Q, Zhang L, Zan T, et al (2020) Wheat RING E3 ubiquitin ligase TaDIS1 degrade TaSTP via the 26S proteasome pathway. Plant Sci 296:110494. https://doi.org/10.1016/j.plantsci.2020.110494

34. Melcher K, Ng LM, Zhou XE, et al (2009) A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. Nature 462:602–608. https://doi.org/10.1038/nature08613

35. Ma Y, Szostkiewicz I, Korte A, et al (2009) Regulators of PP2C Phosphatase Activity Function as Abscisic Acid Sensors. Science 324:1064–1068. https://doi.org/10.1126/science.1172408

36. Song L, Wang J, Jia H, et al (2020) Identification and functional characterization of NbMLP28, a novel MLP- like protein 28 enhancing *Potato virus Y* resistance in *Nicotiana benthamiana*. BMC Microbiol 20:1–14. https://doi.org/10.1186/s12866-020-01725-7

37. Liu J-J, Ekramoddoullah AKM (2006) The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. Physiol Mol Plant Pathol 68:3–13. https://doi.org/10.1016/j.pmpp.2006.06.004

38. Choi SH, Hong MK, Kim HJ, et al (2015) Structure of ginseng major latex-like protein 151 and its proposed lysophosphatidic acid-binding mechanism. Acta Crystallogr Sect D Biol Crystallogr 71:1039–1050. https://doi.org/10.1107/S139900471500259X

39. Pasternak O, Biesiadka J, Dolot R, et al (2005) Structure of a yellow lupin pathogenesis-related PR-10 protein belonging to a novel subclass. Acta Crystallogr Sect D Biol Crystallogr 61:99–107. https://doi.org/10.1107/S0907444904028173

447 **Figure legends**

448 **Fig. 1**. Identification of major latex-like proteins from *Cucurbita pepo*.

449 **a**. Phylogenetic tree of major latex-like proteins of *Arabidopsis thaliana* and *Cucurbita pepo*.

450 Amino acid sequences of 25 AtMLPs and 21 CpMLPs were aligned, and a phylogenetic tree was

451 constructed by applying the neighbor-joining method using MEGA 7.0. Numbers show bootstrap

452 confidence values from 50 replicates. AtMLPs and CpMLPs are divided into seven subfamilies. Red

453 and green letters indicate AtMLPs and CpMLPs, respectively.

454 **b**. Subcellular localization of major latex-like proteins of *Cucurbita pepo*.

455 The subcellular localization of CpMLPs was investigated using PSORT. 'Not applicable' and low and

456 high values are indicated by gray, white, and orange, respectively. Chlo, chloroplast; Cysk,

457 cytoskeleton; Cysk nucl, cytoskeleton and nucleus; Cyto, cytoplasm; Cyto E. R., cytoplasmic

458 endoplasmic reticulum; E. R., endoplasmic reticulum; Extr, extracellular; Golg, Golgi apparatus; Mito,

459 mitochondrion; Nucl, nucleus; Pero, peroxisome; Plas, plasma membrane.

460

461 **Fig. 2**. Gene and protein structures of major latex-like proteins from *Cucurbita pepo*.

462 **a**. Exon-intron structures of MLP genes from *C. pepo*. Genes with introns and coding sequences were

463 extracted from the *C. pepo* database and visualized using Gene Structure Display Server. CDS, coding

464 sequence; UTR, untranslated region.

465 **b**. Conserved domains of MLPs from *C. pepo*. The domains in amino acid sequences of 21 CpMLPs

466 were analyzed using NCBI Conserved Domain Search. Rounded green rectangles indicate the Bet v 1

467 domain.

468 **c**. Conserved motifs of MLPs from *C. pepo*. The motifs in amino acid sequences of 21 CpMLPs were

469 analyzed using MEME version 5.4.1. Rectangles of each color show conserved motifs. The canonical

470 motifs of MLPs with a low *p*-value were not detected in CpMLP17.

Fig. 3. Expression of CpMLP genes in organs of *Cucurbita pepo* subspecies *ovifera* and *pepo*.

a. *Cis*-acting regulatory elements in the promoter region of *major latex-like protein* genes from *Cucurbita pepo*.The 1-kbp-region upstream of the translation starting position in 21 CpMLP genes was extracted, and *cis*-acting regulatory elements associated with plant hormones and root-specific expression were identified using the PLACE database. ABRE, abscisic acid-responsive element; AuxRE, auxin-responsive element; CKRE, cytokinin-responsive element; ETRE, ethylene-responsive element; GARE, gibberellin-responsive element; JARE, jasmonic acid-responsive element; SARE, salicylic acid-responsive element.

b. RNA sequencing analysis of *major latex-like protein* genes from *Cucurbita pepo* ssp. *ovifera* and *pepo*. The *C. pepo* ssp. *ovifera* cultivars ('Patty Green' [PG] and 'Starship' [ST]) and ssp. *pepo* cultivars ('Gold Rush' [GR] and 'Magda' [MG]) were cultivated under a 16/8 h light/dark cycle at 26 °C for 40 days. The total RNA was extracted from roots, and RNA sequencing was performed. Low and high values are indicated as beige and orange, respectively. As there were no genes consistent with *CpMLP3* and *CpMLP11*, these genes' columns are shown in gray.

c. Expression analysis of *major latex-like protein* genes from *C. pepo* ssp. *ovifera* and *pepo*. The *C. pepo* ssp. *ovifera* cultivar PG and ssp. *pepo* cultivar GR were cultivated under a 16/8 h light/dark cycle at 26 °C for 13 days. Total RNA was extracted from leaves and roots, and qRT-PCR was performed. Primer sequences are listed in Suppl. table S2.

Supplementary fig. 1. 3D structures of major latex-like proteins from *Cucurbita pepo*.

Structures predicted from amino acid sequences of 21 CpMLPs were constructed using AlphaFold2. The internal cavity is shown as filled green circles, and blue and red parts indicate hydrophilicity and

494 hydrophobicity, respectively. Because CpMLP13 has an uncharacterized structure, the structure was

495 not constructed by AlphaFold2.

Table 1. CpMLPs identified in the *Cucurbita pepo* genome.

Gene	Gene ID	Amino acid	Location	Molecular weight (Da)	Subfamily[a]
CpMLP1	Cp4.1LG00g07060.1	159	Chr00: 23796116 .. 23797001 (+)	17574.13	I
CpMLP2	Cp4.1LG00g11610.1	160	Chr00: 34667181 .. 34668029 (-)	17574.00	I
CpMLP3	Cp4.1LG01g04090.1	155	Chr01: 1465294 .. 1467541 (-)	17312.57	II
CpMLP4	Cp4.1LG01g05720.1	151	Chr01: 525051 .. 528639 (+)	17470.33	VI
CpMLP5	Cp4.1LG01g07720.1	629	Chr01: 5254154 .. 5260803 (-)	70387.44	VI
CpMLP6	Cp4.1LG01g07730.1	104	Chr01: 5246964 .. 5247823 (-)	11787.38	VI
CpMLP7	Cp4.1LG01g07740.1	151	Chr01: 5269144 .. 5269890 (-)	17489.10	VI
CpMLP8	Cp4.1LG01g07750.1	151	Chr01: 5263963 .. 5264687 (-)	17526.90	VI
CpMLP9	Cp4.1LG01g07760.1	151	Chr01: 5266793 .. 5267508 (-)	17492.97	VI
CpMLP10	Cp4.1LG01g07770.1	109	Chr01: 5249000 .. 5249462 (-)	12496.73	VI
CpMLP11	Cp4.1LG02g16760.1	445	Chr02: 12551609 .. 12566618 (-)	50292.12	IV
CpMLP12	Cp4.1LG03g05870.1	162	Chr03: 4436896 .. 4438387 (-)	18629.32	IV
CpMLP13	Cp4.1LG03g05890.1	842	Chr03: 4442904 .. 4474190 (-)	108182.63	IV
CpMLP14	Cp4.1LG03g05900.1	150	Chr03: 4433116 .. 4433645 (-)	17051.51	IV
CpMLP15	Cp4.1LG03g05930.1	205	Chr03: 4439463 .. 4442277 (+)	23552.10	IV
CpMLP16	Cp4.1LG12g02820.1	176	Chr12: 1980900 .. 1982416 (+)	19345.30	I
CpMLP17	Cp4.1LG14g00140.1	160	Chr14: 4587872 .. 4588422 (-)	18115.94	I
CpMLP18	Cp4.1LG16g01870.1	159	Chr16: 4037155 .. 4038040 (-)	17574.13	I
CpMLP19	Cp4.1LG16g01890.1	294	Chr16: 4081793 .. 4092287 (-)	32341.80	I
CpMLP20	Cp4.1LG16g01910.1	160	Chr16: 4077476 .. 4078324 (-)	17597.09	I
CpMLP21	Cp4.1LG17g00490.1	142	Chr17: 293620 .. 294354 (+)	15392.50	I

497 [a], Subfamily is described in Fig. 1a.

498 Table 2. Cavity volume of CpMLPs.

MLP	Volume (Å^3)
CpMLP1	68.210
CpMLP2	407.627
CpMLP3	127.044
CpMLP4	64.615
CpMLP5	28575.611
CpMLP6	1057.709
CpMLP7	100.230
CpMLP8	74.714
CpMLP9	76.373
CpMLP10	224.345
CpMLP11	4854.621
CpMLP12	376.589
CpMLP14	77.109
CpMLP15	1446.632
CpMLP16	711.831
CpMLP17	239.963
CpMLP18	68.210
CpMLP19	678.856
CpMLP20	461.503
CpMLP21	1053.828
MLP-PG1	57.687
MLP-GR3	98.010

499

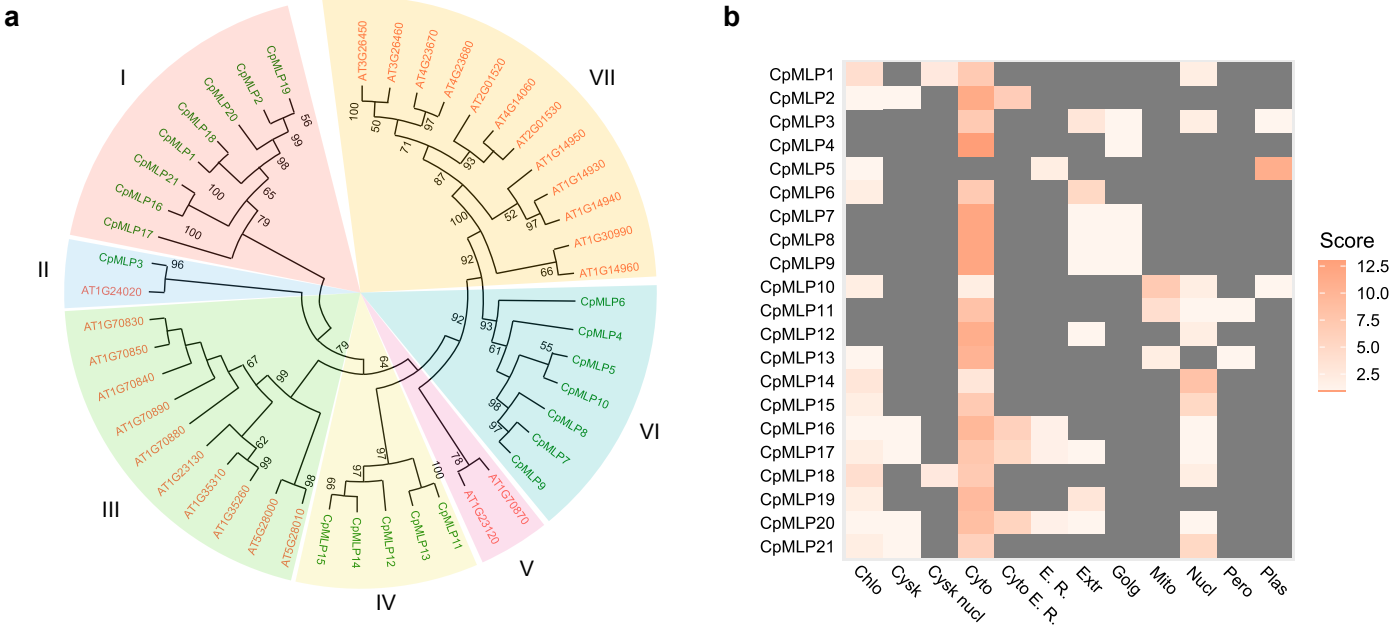

a
I
II
III
IV
V
VI
VII
CpMLP19
CpMLP2
CpMLP20
CpMLP18
CpMLP1
CpMLP21
CpMLP16
CpMLP17
CpMLP3
AT1G24020
AT1G70830
AT1G70850
AT1G70840
AT1G70890
AT1G70880
AT1G23130
AT1G35310
AT1G35280
AT3G28000
AT5G28010
CpMLP15
CpMLP14
CpMLP12
CpMLP13
CpMLP11
AT1G70970
AT1G23120
CpMLP9
CpMLP7
CpMLP8
CpMLP10
CpMLP5
CpMLP4
CpMLP6
AT1G14960
AT1G30990
AT1G14940
AT1G14930
AT1G14950
AT2G01530
AT4G14060
AT2G01520
AT4G23680
AT4G23670
AT3G26460
AT3G26450
b
CpMLP1
CpMLP2
CpMLP3
CpMLP4
CpMLP5
CpMLP6
CpMLP7
CpMLP8
CpMLP9
CpMLP10
CpMLP11
CpMLP12
CpMLP13
CpMLP14
CpMLP15
CpMLP16
CpMLP17
CpMLP18
CpMLP19
CpMLP20
CpMLP21
Chlo
Cysk
Cysk nucl
Cyto
Cyto E. R.
E. R.
Extr
Golg
Mito
Nucl
Pero
Plas
Score
12.5
10.0
7.5
5.0
2.5

Figure 1

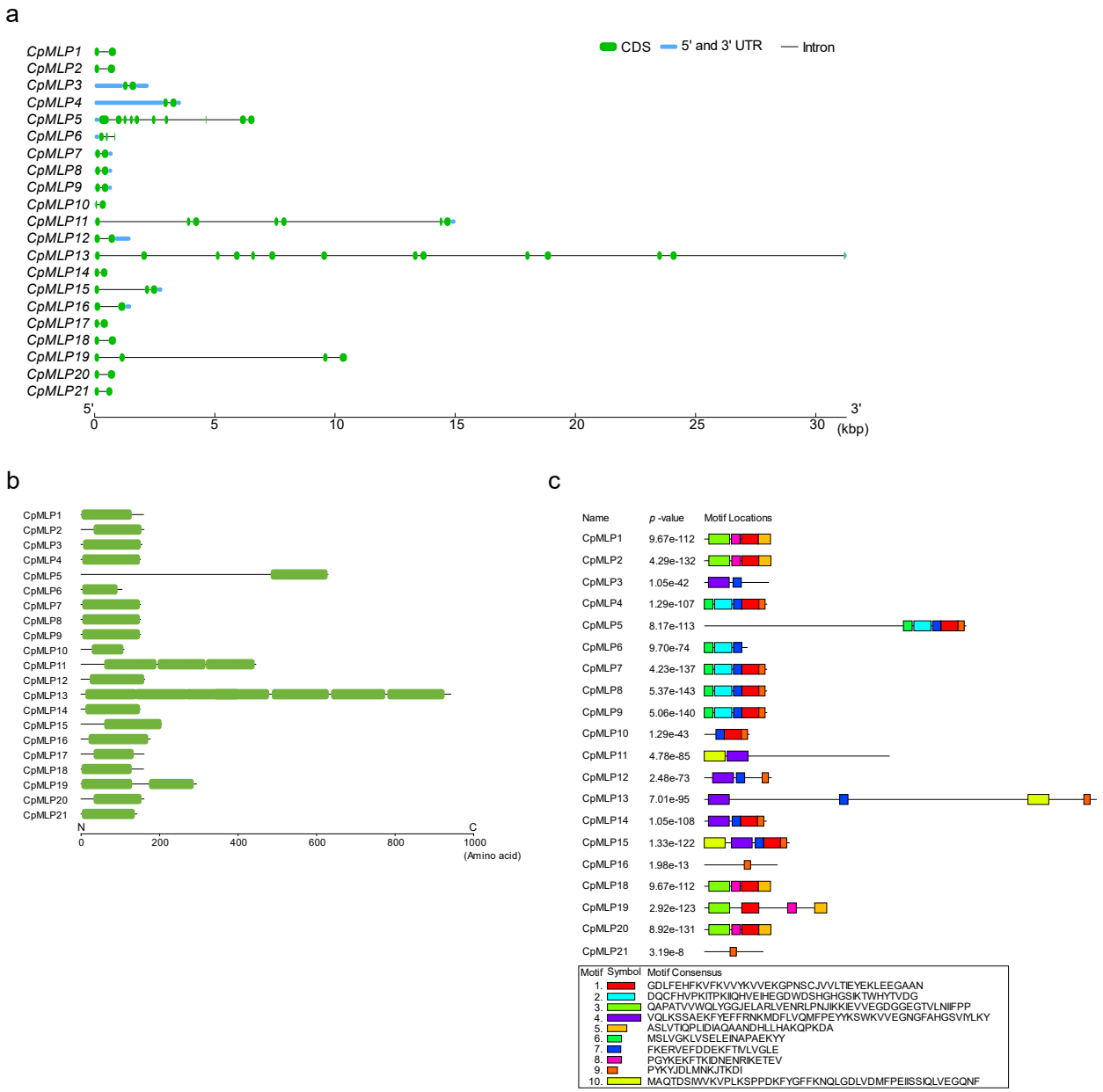

a
CDS
5' and 3' UTR
Intron
CpMLP1
CpMLP2
CpMLP3
CpMLP4
CpMLP5
CpMLP6
CpMLP7
CpMLP8
CpMLP9
CpMLP10
CpMLP11
CpMLP12
CpMLP13
CpMLP14
CpMLP15
CpMLP16
CpMLP17
CpMLP18
CpMLP19
CpMLP20
CpMLP21
5'
3'
0
5
10
15
20
25
30
(kbp)
b
N
C
0
200
400
600
800
1000
(Amino acid)
c
Name
p -value
Motif Locations
CpMLP1 9.67e-112
CpMLP2 4.29e-132
CpMLP3 1.05e-42
CpMLP4 1.29e-107
CpMLP5 8.17e-113
CpMLP6 9.70e-74
CpMLP7 4.23e-137
CpMLP8 5.37e-143
CpMLP9 5.06e-140
CpMLP10 1.29e-43
CpMLP11 4.78e-85
CpMLP12 2.48e-73
CpMLP13 7.01e-95
CpMLP14 1.05e-108
CpMLP15 1.33e-122
CpMLP16 1.98e-13
CpMLP18 9.67e-112
CpMLP19 2.92e-123
CpMLP20 8.92e-131
CpMLP21 3.19e-8
Motif Symbol
Motif Consensus
1. GDLFEHFKVFKVVYKVVEKGPNSCJVVLTIEYEKLEEGAAN
2. DQCFHVPKITPKIIQHVEIHEGDWDSHGHGSIKTWHYTVDG
3. QAPATVVWQLYGGJELARLVENRLPNJIKKIEVVEGDGGEGTVLNIIFPP
4. VQLKSSAEKFYEFFRNKMDFLVQMFPEYYKSWKVVEGNGFAHGSVIYLKY
5. ASLVTIQPLIDIAQAANDHLLHAKQPKDA
6. MSLVGKLVSELEINAPAEKYY
7. FKERVEFDDEKFTIVLVGLE
8. PGYKEKFTKIDNENRIKETEV
9. PYKYJDLMNKJTKDI
10. MAQTDSIWVKVPLKSPPDKFYGFFKNQLGDLVDMFPEIISSIQLVEGQNF

Figure 2

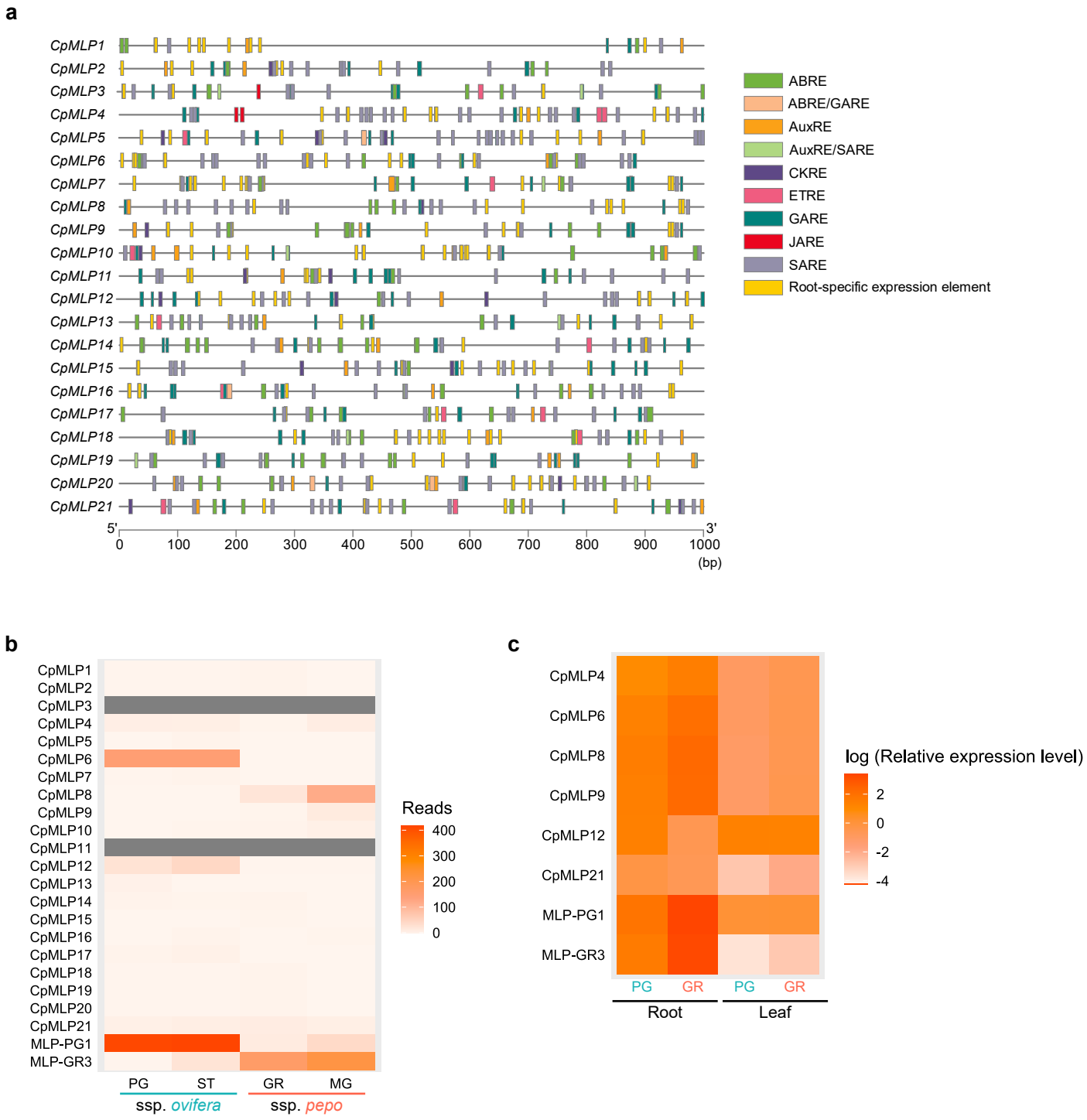
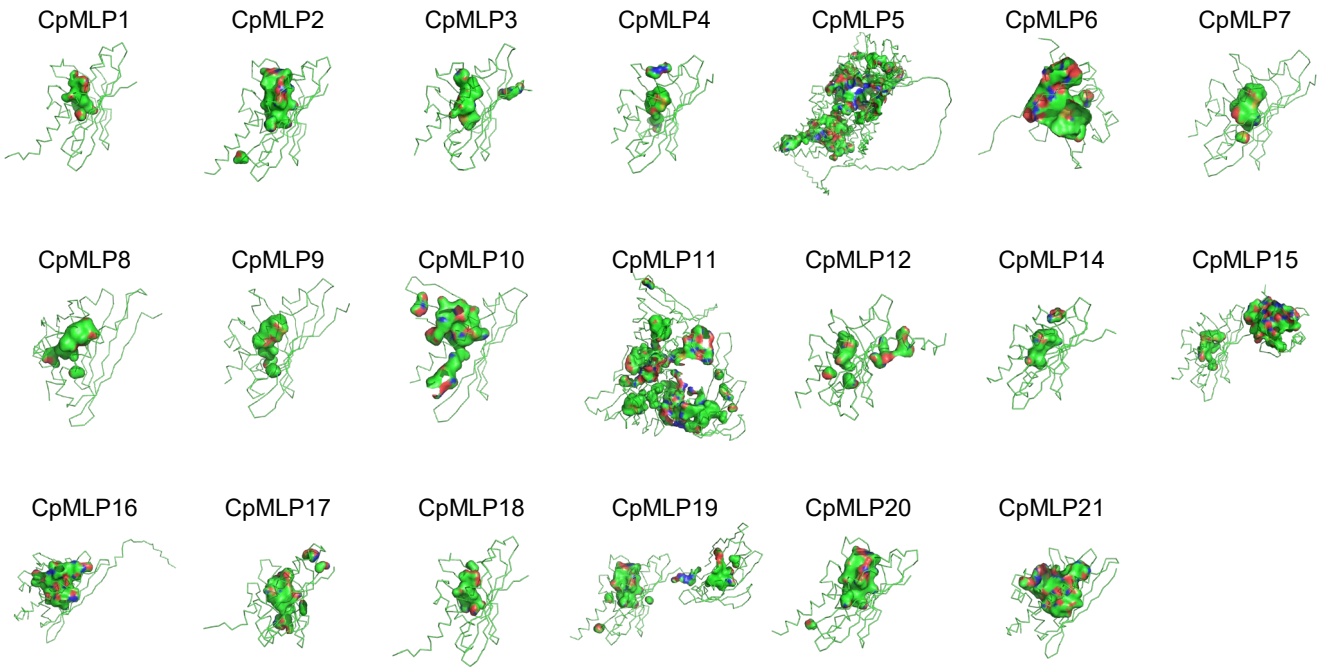

a
CpMLP1
CpMLP2
CpMLP3
CpMLP4
CpMLP5
CpMLP6
CpMLP7
CpMLP8
CpMLP9
CpMLP10
CpMLP11
CpMLP12
CpMLP13
CpMLP14
CpMLP15
CpMLP16
CpMLP17
CpMLP18
CpMLP19
CpMLP20
CpMLP21
5'
3'
0
100
200
300
400
500
600
700
800
900
1000
(bp)
ABRE
ABRE/GARE
AuxRE
AuxRE/SARE
CKRE
ETRE
GARE
JARE
SARE
Root-specific expression element
b
Reads
400
300
200
100
0
PG
ST
GR
MG
ssp. ovifera
ssp. pepo
c
log (Relative expression level)
2
0
-2
-4
CpMLP4
CpMLP6
CpMLP8
CpMLP9
CpMLP12
CpMLP21
MLP-PG1
MLP-GR3
PG
GR
PG
GR
Root
Leaf

Figure 3



Supplementary figure 1

Supplementary table 1. Amino acid sequences of CpMLPs.

MLP	Sequence
CpMLP1	MVTIISDQTEIPAPAAKVWALYGTIHFADFLQLHLPNIINNVELLEGDGGQG TLVLVTFAPDLGGMRYKEKFVKIDNEQRIKIAEMVEGGYLDLGFTVYRFCF EIIEKDEESCIVKSSVEYELKEEAAANVSLASVQPLIAIAQAAKSYFLNAQQ PTDA
CpMLP2	MLGQLSHEAAIQAPATVVWQLYGGLELARLIENRLPNLIKKIEVVEGDGGE GTVLNIIFPPGLGGAPGYKEKFTKIDNENRIKETEVVEGGFLDIGFTLYRVR LKIVENGDDSCIVESTIEYDIKEEDAANASLVTIQPLIDIAQAANDHLLHNKQ HKNV
CpMLP3	MASDGTLNVEVDVKSVAPKFWNSMRDSTIIFPKAFPHDYKSIEVLEGDGK AVGSVRLITYSEGSSLVKDSKERIEAVDEEAMTVSYSVIEGDLLKYYKSFK GHIGVIPKEDGSGSKVKWSCEFEKASEEVPDPHVIKDFVVKNFLELDDYV LQQP
CpMLP4	MALAGKLVSEVKINVAAEKYYKIWKHEVSHVPKICPKYIQKVEVHEGDWD SHGHGSVKIWHYSIDGKAGFFKERVEFDDKNMAMLLVGLDGDMFEHYKS FKATYKVVPKGPNHCLVVMILKYEKLRADCPSPYKYIDLMNDLTKSLESYL Q
CpMLP5	MGASHQWSSSLQTAPIKLKSSIPLTSPSNFIFYYCKRSRVNYSSTSRCAVC AHNSNLPRPKSTNSDARISKSVVLGDCQGHELVRISSTSIRRRKSVILSLV SLFDKRSLWRRIFFASKKVRSIILLNIVTIVYASSIPVVKEVEELVDPATFNAV RFAITAIPFVPLVLYKWDDVETRNAGIELGFWVSLGYLMQAFGLLTSDAGR ASFISMLTVLVVPILDGVLGAVVPARTWFGVLMSVIGVAMLESSGSPPCVG DLLNFLSAIFFGVHMLRTEHISRRTEKDKLVPLLAYEVCVVSILSMLWYFIW RWIDGTETISESWNWKTYSDWVFMFPWVPALYTGLLSTGFCLWLEMGA MCDVSATETAVIYSLEPVWGGSFAWFLLGERWGLSGWIGAALVLGGSLTV QILSSSATKSCKDDRSKEVHDVLGSADKRSLSTSPIVLTRVLAPRLRENVE GKPLGSVICKPISNRVSETMSLRGKFVSELELNAAAHKYYKLFKHQVSHIP NISPGIFKNVEVHEGDWDTHGHGSIKIWNYNIDGKDEVFKEQVEFDDEKL SVTLIGLEGDVFEHYKTFKGIYQVVPKGPEHCLAVLTLEYEKLDDGSPYPY KYLDLMNNLTRDIESHLK
CpMLP6	MSLAGKLVSEIEINVAAEKYYKVFKDQPFNVPNISPKLVQQVELDEGDWD NHGHGSVKTWKYTVDGKPEVFKEKAEFDDEKFTIIMNGLQGDVRVKINK GYPGV

CpMLP7 MSLVGKLVSELEINIPAEKYYKVFKDQCFHVPKITPKIIQHVEIHDGDWDSH
DHGSIKTWHYTVDGKSEVFKERVEFHDEKFMVVLVGLEGDLFNHYKTFK
PVCQVVPKGPSHCLAVLTIEYEKLDDGSPYPFKYIDLMNGITKDIESHLK

CpMLP8 MSLVGKLVSELEINAPAEKYYKVFKDQCFHVPNITPKFIQHVEIHEGDWDS
HDHGSIKTWHYTVDGKSEVFKERVEFHDEKFTIVLVGLEGDVFNHYKTFK
PVYQVVPKGPSHCLAVLTIEYEKLDDGSPYPYQYIDLMNGITKDIESHLK

CpMLP9 MSLVGKLVSELEINAPAEKYYKVFKDQCFHVPKITPKIIQHVEIHDGDWDS
HDHGSIKTWHYTVDGKSEIFKERVEFHDEKFTVVLVGLEGDVFNHYKTFK
PVYQVVPKGPSHCLAVLTIEYEKLDDGSPYPYKYIDLMNGITKDIESHLK

CpMLP10 MSLKFITATGRIIVMARSKSGITLLAEELKERVEFDDKNLVVCMIGLEGDVF
EHYKVFKAIFKFVPKGPNRSAVILILEYEKLHDGPPYPHKYHDAMHKLAKDI
ESHLK

CpMLP11 MSQTDSIWAKLPLKSPPDTFYGFFKNQVGDFVDMFPEYISSIQLAEGENF
APDSVMQFKYSLEKYYGFFRNHMGDMVNLLPQYFSSIQLVEGANFSPDC
IIQFKYSLGGGSLSAKVKIKAVDDAKKLLAYNVIEGDVLKHYKVFEVRMEVV
NGGTSKGGGGSFAKWSVVFEKANENVAAPEDYLEWFVKISKGFPVKSPP
DKFYGFYRNHVGDLIDLFPQYFSSIQFVEGEKYSPDSVIRFNYRFANIKIKA
VDDVKKSLVYKVIEGDILKHYKVFELRIEAVNGGISKGGGGSFAKWSIVFE
KANENVGAPQGYLEWHKMHHLPQIFSKNLHSFEFLEGNDFTPGSLMHW
SYDIVGPAKMKAKVADVDEENKSITYEAVEGDILSQYTLLRSKFRAYDDVE
NGGAIVNWSFEFEKANENIPSPEAYLEFVSKISIGLDAYLAVN

CpMLP12 MDEHILKYLKMAQISNISHQLQLKCSGEQFYEFYRNKMDRLTQMFPKKLL
GYKIVEGNGFAHGSVVYWKYELGCILEAKQKLHMDDKNKAITLEFIEGDLF
KEYEMIAVKGEVSDGGSNGISSVKWSVEYVKANEDVDPPHNYLQFALEL
AKGVDAYLCNNN

CpMLP13 MSQIESIWGKVQLKSSPEKFFGFFRNHMGDLVHMFPDHFQSFHFVEGQN
FDDGSVVHWKYHLGIPEAVKIKMKNRDEARTIIYEVVEGDALKHYKVFRAK
LETVSGGLNKVGGSFAKWTIEYEKAHENVPSPETYMELALKLKSSPEKFY
GFFRNHMGELVHMFPDHFQSFHFLEGQNFDDGSVVQWKYHLGFPEAAK
VRMRVMDEARTIIYEVVEGDALKHYKAFRVKLETVSGDLNKVGANFAKWT
IEYEKAHQNVASPETYLELALQLKSSPEKFYGFFRNHMGELVHMFPDHFQ
SFHFLEGQNFDDGSVVQWKYHLGFPEAAKVRMRVMDEARTIIYEVVEGD
ALKHYKAFRVKLETVSGDLNKVGANFAKWTIEYEKAHQNVASPETYLELA
LQVTKGFPEAAKVRMRVMDEARTIIYEVVEGDALKHYKAFRVKLETVSGD
LNKVGANFAKWTIEYEKAHQNVASPETYLELALQKMGKSDSIWAKIDLKS
SPEKFYGFFRNHLGDLVDLFPENYKSIQLVEGQHFSGGNVVLFKFQFGFG

HQLRVEKWAIRAVDDVKKYIIYEAVEGDVLKQFKVLRVKVEAVHGGSTKV
GGGNFTKWTVEFEKANQNVASPQNYLELFVKISKGTMGKSDSIWAKVDL
KSSHEKFYGFFRNHLGDLVDLFPENFNSIQLVEGQHFDRGSLVLRHEHRV
EKWVIRAVDDVKKYIVYEAVEGEALKQFKVLRAKVEAVHGGSTKVGGGNF
TKLTIEFEKANENVASPEIYLELFVKIAKGKMVQTDSIWVKVDLKSSPEKVY
GFFRNHLGDLVDLFPETYQSIQLVEGQHFSSGSVVQFKFQFGDELRAEK
WAIRVVDDVKKYIIYEAVEGDPLKEFKVLRAKFEVVNGGLSKVRRGNFTK
WTVEFEKANQNVASPQNYLELFVKISKGILVDVESVQKLCGYNVT

CpMLP14 MAQIAKVSQKVQLRSSGHKFYELLKNKMDFVFQMFPEVYKSWKVLEGN
GLAHGSIIYLKYDVDGLSEAKERLAIDDANKSITFECLEGDLFRDFEVFKLK
IEVVENGSNGCSSNWSIEYVKANEDVAPPHNYLIIAAKISKGIDDYLCKN

CpMLP15 MAQNAKISRQVQLKCCGHKFYELFKNKMGCVFQMFPEICSSWKVLEGN
EFAHVRVIHVKYVVSQEVQLRSSAPKFYEFLKNKMDFVFQMFPEIYKSWK
VVEGNGYAHGSVIQLKYNVDGPSEVKERLTIDDANKSLTFECVEGDLLRD
FEVFKMKIEVVENGSNGSSANWSIEFVKANEDVATPHNYLLCVAKVSKGI
DDYLCKN

CpMLP16 ISFSLFLLLKLPSCEIAMGVFTYENEVTTVIPPAKFFKAFILDADRLYPKIVPH
QPKTEVVEGDGGPGTIKIITFSHGGQVKSIKHRLDVVDEKSLTYKYTVLEG
ELLSDNIDQISKELKVTAGPDGGSILKSVSIYHTKGDHQIDEQKLKIGEEKG
LGLLKAAEAYLLANPAEYN

CpMLP17 MVTKEAKAEAKLGVEIETLWKALAKDLRFIIPQLMPDTVEKIELLHGDGGV
GSILLFHLVHKEEAMRSQKERIVEVDETRHELVIQVLEGNVLKRGFSSFKT
TFKLSSLSEKESLVDIKVAYETEKDGEDEQARMDAIATAPPLYFFQLLEKFL
LPTSNT

CpMLP18 MVTIISDQTEIPAPAAKVWALYGTIHFADFLQLHLPNIINNVELLEGDGGQG
TLVLVTFAPDLGGMRYKEKFVKIDNEQRIKIAEMVEGGYLDLGFTVYRFCF
EIIEKDEESCIVKSSVEYELKEEAAANVSLASVQPLIAIAQAAKSYFLNAQQ
PTDA

CpMLP19 MLGKLSHETVIQAPATVAWQLYGGLELARLVENRFSNLIQKIEVVEGDGGE
GTVLNLIFPPGVGRFSSFKEKFTRIDNENRIKETEIVEGGFLDIGFTLYRVC
LKIVENGDDSCIVESTIEYEIKEEAAANASLMLGQLSHEAAIQAPATVAWQL
YGGLELARLVENRLSNLIQKIEVVEGNGGEGTVLNLIFLPGLGGAPSYKEK
FTKIDNENRIKETEVVEGGFLDIGFTLYRVRLKIVENGDDSCIVESTIEYEIK
EEAAANASLVTLQPLIDIAQAANDHLLHYKQLKDA

CpMLP20 MLGHLSHEAVIQAPATVVWKLYGGLELARLVENRLPNLIKKIEVVEGDGGE
GTVLNIIFPPGLGGAPGYKEKFTKIDNENRIKETEVVEGGFLDIGFTLYRVR

	LKIVENGDDSCIVESTIEYDIKEEDAANASLVTIQPLIDIAQAANDHLLHNKQ HKNV
CpMLP21	MGVFTYENEVASVIPPEKFFKAFILGADQLYPKIVPNQPQSVLEGDGGPG TIKTISFSVASRTYKYTVLEGELLSDAIDKISKEIKVVEGPSGGSILKSTSVY HTKGDHQIDEEKLKSGEQKGLALLKAAEAYLLANPNEFN

2

3 Supplementary table 2. List of primers used for qRT-PCR.

Primer	Sequence
CpActin-s	5′-TCCAGGCCGTTTTATCTC-3′
CpActin-as	5′-CAGAATCCAACACAATACCTGT-3′
CpMLP4-s	5′-TGGCCATGGCTCAGTCAAGA-3′
CpMLP4-as	5′-CCAATCCAACCAAAAGCATTGCC-3′
CpMLP6-s	5′-TGGTCGGGAAACTGGTGAGC-3′
CpMLP6-as	5′-TGAATCTCAACGTGCTGGATGA-3′
CpMLP8-s	5′-AGCGAGTGGAATTTCACGACG-3′
CpMLP8-as	5′-ATGGCTAGGACCCTTTGGCA-3′
CpMLP9-s	5′-CCTGGCATTACACAGTTGATGGG-3′
CpMLP9-as	5′-GGCACAACCTGATATACCGGC-3′
CpMLP12-s	5′-TCGGGTGCATACTAGAGGCA-3′
CpMLP12-as	5′-TGCCATTGCTCCCACCATCA-3′
CpMLP21-s	5′-ACGAAAACGAAGTGGCGTCG-3′
CpMLP21-as	5′-ACGGATTGAGGCTGATTTGGC-3′
MLP-PG1 -s	5′-ATTCAAAGTGCTAAGAGCAAAAT-3′
MLP-PG1 -as	5′-CCTTTTCAAACTCAACAGTCCA-3′
MLP-GR3-s	5′-AATTCAAAGTGCTTAGAGCAAAGG-3′
MLP-GR3 -as	5′-TGCCTTTTCAAACTCAATAGTCAA-3′

4