MePHD1.2 affects the synthesis of cyanogenic glycosides by regulating the transcription of *MeCYP79D2* in cassava

Mengtao Li[1,2], Xiao Zhao[2,3], Yuanchao Li[2,4], Xiaoye Zhao[5], Weitao Mai[1,2], Yajun Li[2], Qibing Liang[5], Qingchun Yin[6], Wenquan Wang[1*], Jinping Liu[1*], Xin Chen[1,2,7*].

[1]Institute of Tropical Agriculture and Forestry, Hainan University, Haikou, 570228, China/National Key Laboratory for Tropical Crop Breeding, Sanya 572025, China.

[2]Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, 571101, China.

[3]College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, 430070, China.

[4]State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources/College of Life Science and Technology, Guangxi University, Nanning, 530004, China.

[5]Haikou Branch of Hainan Food and Drug Inspection Institute, Haikou 570311, China.

[6]Hainan Institute for Food Control, Key Laboratory of Tropical Fruits and Vegetables Quality and Safety for State Market Regulation, Haikou, 570314, China.

[7]Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences, Sanya, 572025, China.

Authors for correspondence:

Wenquan Wang

E-mail: wangwenquan@hainanu.edu.cn

Jinping Liu

E-mail: liu3305602@163.com

Xin Chen

E-mail: chenxin@itbb.org.cn

Summary

The high content of cyanogenic glycosides (CG) in cassava storage tubers seriously affects human food safety. CG play crucial roles in plant growth and development and can protect cassava leaves from being masticated by herbivorous predators.

Nevertheless, the regulatory mechanism of CG biosynthesis, which results in a low CG content in storage tubers and high CG content in leaves, remains poorly understood.

Here, yeast one-hybrid assay was performed using a mixed cDNA library of cassava storage roots and leaves as prey and the promoter of *MeCYP79D2* as bait. MeCYP79D2, a cytochrome P450 protein, is the rate-limiting enzyme for CG synthesis in cassava. From this information, a candidate regulator of *MeCYP79D2*, that is, transcription factor MePHD1.2, was selected.

MePHD1.2, which is located in the nucleus and exhibits a transcription inhibitory activity, can directly bind to PD2 segment in the promoter of *MeCYP79D2*, which results in its repressed expression. In cassava, the transcriptional activity of *MeCYP79D2* was considerably enhanced in *mephd1.2* lines, which caused an increase in the contents of linamarin and lotaustralin.

Our findings unveil a novel regulatory module governing CG biosynthesis, wherein mutation of *MePHD1.2* attenuates its transcription inhibition on *MeCYP79D2* and boosts CGs biosynthesis in cassava.

Key words: Cassava, Cyanogenic glycosides, Cytochrome P450, MePHD1.2

Introduction

Cassava (*Manihot esculenta* Crantz) belongs to the three major tuberous crops in the world, and it is widely cultivated in tropical regions. Its root is considered a staple food for nearly one billion people living in tropical and subtropical regions (Panghal *et al.*, 2019). Cassava, whose leaves contain over 20% crude protein, exhibits an enormous potential for applications in animal feed and human food production. However, its aerial parts features high concentrations of cyanogenic glycosides (CG), which limit its utilisation in animal feed or human consumption (Latif & Müller, 2015).

CGs comprise plant products or secondary metabolites derived from amino acids (AAs) with oximes and cyanohydrins (α-hydroxynitriles) as key intermediates; these compounds are widely distributed in all tissues of cassava, mainly synthesised in the leaves and transported to the roots (Møller, 2010; Kannangara *et al.*, 2011; Gleadow and Møller, 2014). CGs can help in the plant defence against external encroachment

and serve as deterrents to herbivores (Zagrobelnya *et al*., 2004; Ballhorn *et al.*, 2016). CGs, specifically linamarin and lotaustratalin, exist in all tissues of cassava. When hydrolysed, these compounds can release hydrogen cyanide (HCN) (Santana *et al.*, 2002; Ogbonna *et al.*, 2021). In cassava, CGs also function as storage compounds of reduced nitrogen and carbon and quenchers of reactive oxygen species (Frederik *et al*., 2018). However, long-term consumption of cassava storage roots with high CG contents can cause nausea, diarrhoea and other symptoms (Banea-Mayambu *et al.*, 1997; Lebot, 2019). Therefore, cultivation of new cassava varieties with high CG amounts in leaves and low CG levels in storage roots is one of the currently important breeding goals.

Cytochrome P450s are characteristic enzymes that are ubiquitously present in numerous plants and animal species; they play important roles in various biochemical pathways (Zagrobelny *et al*., 2004). Most cytochrome P450s are associated with the biosynthesis of terpenoids, volatile derived fatty acids, plant hormones (gibberellins) and brassinosteroids (Dudereva *et al*., 2004; Fischbach and Clardy, 2007). In spite of species differences, CYP79s are the primary key enzymes in CG production, and they are considered rate-limiting enzymes of cyanogenesis pathways (Gleadow and Møller, 2014). The synthesis process of CGs involves three steps in cassava (Schmidt *et al.*, 2018). Firstly, the dehydration and decarboxylation of L-valine and L-isoleucine occur via the catalysis of MeCYP79D1 and MeCYP79D2, which results in the production of (Z)-acetaldoxime, respectively. Subsequently, MeCYP71E7 catalyses (Z)-acetaldoxime to produce cyanohydrin, acetone cyanohydrin and 2-hydroxy-2-methylnitrile (Jørgensen *et al.*, 2011). Finally, UDP glucosyltransferases, namely, MeUGT85K4 and MeUGT85K5, catalyse the production of linamarin and lotaustralin, respectively, from acetone cyanol and 2-hydroxy-2-methylnitrile (Fig. 1; Andersen *et al.*, 2000; Kannangara *et al.*, 2011). Hydrolysis of CGs occur via the continuous enzymatic reaction of β-glucosidase and α-hydroxynitrilase, which forms glucose, aldehydes or ketones and toxic hydrogen cyanide when cassava organs are suffering from mechanical damage (McMahon *et al.*, 2022).

When animal gnawing damage the plants of cassava, lima bean, sorghum and other crops, the expression levels of *CYP79Ds* genes are substantially up-regulated, and the synthesis efficiency of CGs and dhurrin increase, which reduces palatability of the

103 plants and prevents further gnawing by animals (Lai *et al.*, 2020; Easson *et al.*, 2021).
104 The simultaneous inhibition of the expressions of *MeCYP79D1* and *MeCYP79D2*
105 through RNAi can greatly reduce CGs contents in cassava leaves and roots (Jørgensen
106 *et al*., 2005). In addition, during the knockout of *MeCYP79D1* and *MeCYP79D2* by
107 gene editing, the dual knockout lines showed no cyanogenic potential, and
108 *MeCYP79D2* knockouts revealed a more drastic reduction in cyanogenic potential
109 than *MeCYP79D1* knockouts and wild types, which indicates that *MeCYP79D2* is a
110 dominant contributor to the biosynthesis of CGs in cassava (Gomez *et al*., 2023).
111
112 The transcription factor LjbHLH can up-regulate the transcriptional activity of
113 *LjCYP79D3* in *Lotus japonicus* through induction of jasmonic acid and promote the
114 synthesis of lotaustralin (Chen *et al.*, 2022). In cucumber, the transcription factor
115 CsBl causes the up-regulation of the expression of *CsBi* to promote cucurbitacine
116 synthesis in leaves. Moreover, the alteration in Cys393Tyr in CsBl results in the
117 reduced synthesis efficiency of cucurbitacine because CsBl cannot bind to the
118 promoter of *CsBi*, which also *CsBi* encodes the enzyme responsible for the catalysis
119 of the first and rate-limiting step in cucurbitacine synthesis pathway (Shang *et al.*,
120 2014). Similarly, the alteration in Leu346Phe caused PdbHLH2 to form an unstable
121 homodimer, as a result, its binding capability to the promoters of *PdCYP79D16* and
122 *PdCYP79AN24* was reduced, the transcription activities of its targets were
123 down-regulated, and the synthesis of amygdalin (a type of CG) in the endosperm of
124 almonds, which led to the generation of sweet almonds, decreased (Sánchez-Pérez *et*
125 *al.*, 2019).
126
127 Plant homeodomain (PHD) refers to a class of zinc finger proteins that contain
128 Cys4-His-Cys3 zinc-binding motifs and is involved in various molecular functions
129 (Ma *et al.*, 2018). The PHD-finger domain can specifically recognise the methylated
130 modification of histone and may serve as an important reader of histone cryptogram
131 (Jain *et al.*, 2020). Early bolting in short days (EBS) of *Arabidopsis* involves two
132 Bromo adjacent homology (BAH)-PHD domains that bind to H3K27me2 and
133 H3K4me3 as a bivalent histone reader to mediate the switching of binding preferences
134 between H3K27me2 and H3K4me3 and modulate flowering time in *Arabidopsis*
135 (Qian *et al.*, 2021). Alfin1-like proteins (ALs), which are members of the PHD-finger
136 family, can read the epigenetic modification of histone, e.g. methylate H3K4me2/me3,

137 repress the transcription activities of *AtbZIP50*, *AtCAX1*, *AtFAO* and *AtWRKY11*
138 genes and intensify the tolerances to drought and salt. ALs commonly bind to the
139 G-rich motif (GTGGAG), and the mutations of its AAs at positions 34 and 35 in AL6
140 caused the loss of ability to bind the G-rich motif (Wei *et al.*, 2015).
141
142 However, available information on the PHD transcription factor involved in the
143 synthesis of CG is limited. In the present study, the promoter of *MeCYP79D2* was
144 cloned to construct a bait vector for yeast one-hybrid assay using a mixed cDNA
145 library as prey, and the transcription factor MePHD1.2 was screened out. MePHD1.2
146 is a nucleus-localised protein, and it can bind to the PD2 segment in the *MeCYP79D2*
147 promoter, down-regulate its transcriptional activity and increase the CG content in the
148 leaves of *mephd1.2* mutant.
149
150 **Materials and Methods**
151 Plant materials and growing conditions
152 A cassava cultivar (SC8) and friable embryogenic callus (FEC) were cultivated in the
153 dark at 28°C in sterile culture rooms, and *in vivo* cassava plants were inoculated in
154 high-transparency silica glass culture bottles containing Murashige and Skoog basal
155 salt (MS) medium. A photoperiod of 16 h/8 h light/dark condition, with a temperature
156 of 28°C, was adopted. The genetically modified *in vitro* cassava materials were
157 planted in an environment with humidity ranging from 60% to 80%. In the case of
158 tobacco, the optimal culture conditions comprised 14 h of light followed by 10 h of
159 darkness at 24°C.
160
161 Isolation and analysis of the promoter of key enzyme genes in the CG biosynthesis
162 pathway
163 In this study, the AM560 reference genome of cassava from Phytozome
164 (https://phytozome-next.jgi.doe.gov/) was used, and genomic DNA was extracted
165 from cassava via the cetyltrimethylammonium bromide (CTAB) method.
166 Subsequently, a 2 kb sequence upstream of the *MeCYP79D2* start codon was
167 amplified (Fig. S1). In addition, cis elements within promoter sequences were
168 predicted using the web tool PlantCARE
169 (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/; Table S1).
170

171 Yeast one-hybrid assay

172 The promoter fragments (-1 to -1505) of *MeCYP79D2* were inserted into the *pAbAi*
173 vector for bait construction. After digestion with *Bst*B I, the *pAbAi-MeCYP79D2* bait
174 vector was used to transform yeast strain Y1HGold and integrated into the yeast
175 genome as reporter strains. The *pAbAi-MeCYP79D2* yeast solution with an optical
176 density $(OD)_{600}$=0.006 was spread evenly on the SD/-Ura medium containing various
177 gradient concentrations of aureobasidin A (AbA, 0–250 ng/mL) to inhibit
178 self-activation of the promoter. A prey vector containing a mixed cDNA library
179 plasmid from cassava storage roots and leaves was introduced to the reporter strains,
180 and growth on the SD/-Ura-Leu (AbA, 300 ng/mL) medium at 30°C for 3–5 days.
181 The positive colonies were identified via polymerase chain reaction (PCR) with AD
182 detection primer pairs (forward:
183 5'-CTATTCGATGATGAAGATACCCCACCAAACCC-3' and reverse:
184 5'-GTGAACTTGCGGGGTTTTTCAGTATCTACGAT-3'; Table S2).

185

186 The point-to-point yeast one-hybrid assay followed the abovementioned steps, and the
187 full-length coding sequence (CDS) of *MePHD1.2* was cloned into the pGADT7 vector.
188 The combination of *pAbAi-P53* and *pGADT7-Rec53* served as a positive control and
189 that of *pAbAi* and *pGADT7-Rec53* as a negative control. The concentration of yeast
190 solution with 0.9% NaCl was diluted to OD_{600}=0.6, and a pipette was used to collect
191 10 μL solution in a 10^{-1} gradient to form a spot on the culture medium. Cultivation
192 was conducted at 30°C for 3–5 days.

193

194 Phylogenetic tree construction

195 The AA sequences of *Arabidopsis* and other species proteins, which are similar to
196 those of MePHD1.2, were downloaded from the National Center for Biotechnology
197 Information (https://blast.ncbi.nlm.nih.gov), and a phylogenetic tree was constructed
198 using the neighbour-joining (NJ) method in MAGA6 software.

199

200 Extraction of total RNA and gene expression analyses

201 Cassava tissue was frozen using liquid nitrogen, and total RNA was extracted using
202 the Plant Total RNA Extraction Kit (Fuji, Chengdu, China). cDNA synthesis was
203 completed using the reverse-transcription reagent MonScript™ RTIII Super Mix with
204 dsDNase (Monad, Suzhou, China). Quantitative real-time PCR (qRT-PCR) was

205 performed using MonAmp™ ChemoHS qPCR Mix (Monad, Suzhou, China) and
206 LightCycler® 480 II (Roche Diagnostics, Germany). Table S2 contains the primer
207 sequences. Each sample underwent three reactions. The relative expression levels of
208 target genes were normalised to that of *Tublin* via the $2^{-\Delta\Delta CT}$ method.
209
210 We collected primary cassava leaf tissues from three individual overexpression (OE)
211 and *cas* transgenic lines (OE-1, OE-2, OE-3, *cas-1*, *cas-2* and *cas-3*), along with a
212 control check (CK) line. Each sample included three independent plants. Then,
213 RNA-seq was performed by GENE DENOVO company (Guangzhou, China). The
214 raw sequence data were deposited to the Genome Sequence Archive (Genomics,
215 Proteomics & Bioinformatics 2021) at the National Genomics Data Center (Nucleic
216 Acids Res 2022), China National Center for Bioinformation, Chinese Academy of
217 Sciences (GSA: CRA014863) (https://ngdc.cncb.ac.cn/gsa).
218
219 Subcellular localisation of MePHD1.2
220 The *pSL1 Plus* vector containing green fluorescent protein (GFP) was used as the
221 skeleton carrier. The CDS of *MePHD1.2* without a stop codon was fused with the GFP,
222 which is driven by the 35S promoter, to obtain *pSL1 Plus::MePHD1.2-GFP*. Then,
223 the protoplasts were isolated from *in vivo* SC8 cassava leaves of plants grown in an
224 artificial-climate incubator. The cassava leaves were digested using Cellulase R10
225 (Yakult Honsha, Tokyo, Japan) and Macerozyme R10 (Yakult Honsha, Tokyo, Japan)
226 to prepare the mesophyll cell protoplasts. Polyethylene glycol-mediated cell fusion
227 was used to introduce the nuclear marker mCherry and fusion constructs of
228 *MePHD1.2-GFP* to the cassava mesophyll cell protoplasts. After 12 h incubation in
229 the dark, the protoplasts were harvested, and GFP fluorescence was examined via
230 confocal microscopy.
231
232 Transactivation activity assay
233 The transcription activation ability was verified using a yeast system, and the
234 full-length CDS of *MePHD1.2* was integrated into the *pGBKT7* vector. Next, the
235 *pGBKT7-MePHD1.2* fusion vector was transformed into Y2HGold yeast strain. We
236 adjusted the OD_{600} of the bacterial liquid to 0.06. The bacterial liquid was
237 subsequently inoculated onto SD/-Trp, SD/-Trp-His, SD/-Trp-Ade and
238 SD/-Trp-His-Ade media. The cultures were incubated at 30°C for 3–5 days.

239

240 Dual luciferase reporter assay was performed to test the transcription activation
241 capability of MePHD1.2. The reconstructed *pGreenII 0800-LUC* vector containing
242 five copies of the GAL4 binding element (5×GAL4) served as the reporter (Wei *et al.*,
243 2017). As the effector, MePHD1.2 was fused with the *GAL4* DNA-binding domain
244 (*GAL4BD*), and an empty *GAL4BD* was used as a control. Tobacco leaves were
245 ground in 1×lysis buffer solution, and the ratio of LUC and REN was determined
246 using the of Dual-Luciferase® Reporter Assay System test kit (E1910, Promega,
247 America). All samples were subjected to more than three repeated measurements.

248

249 Yeast two-hybrid assay

250 *pGBKT7-MePHD1.2* and *pGADT7* vectors were cotransformed intoY2HGold yeast
251 strain, and the bacterial solution was evenly dispersed on the SD/-Trp-Leu medium.
252 PCR validation involved single colonies and used the AD primer pair. The positive
253 strains were selected for activation. Self-activation validation was carried out on the
254 SD/-Trp-Leu-His-Ade medium with varying concentrations of AbA. Meanwhile,
255 *pGADT7-MePHD1.2* and *pGADT7-MePHD1.1* were cotransformed with the
256 *pGBKT7-MePHD1.2* plasmid into Y2HGold. Protein–protein interactions were
257 determined via colony transfer to the SD/-Trp-Leu-His-Ade medium.

258

259 Dual-Luciferase reporter assay

260 The promoter fragments (-1 to -1505) of *MeCYP79D2* were inserted into the *pGreenII*
261 *0800-LUC* vector as the reporter. The full-length CDS of *MePHD1.2* was cloned into
262 the *pGreenII 62-SK* vector as effector. Along with the *pSoup* helper plasmid,
263 recombinant constructs were introduced to *Agrobacterium* GV3101, followed by their
264 injection into tobacco leaves. Empty *pGreenII 0800* and *pGreenII 62-SK* were used
265 collectively as negative controls. Approximately 48 h later, the tobacco leaves were
266 harvested for measurement of luciferase activity using the same method as the
267 transactivation activity assay.

268

269 Electrophoretic mobility shift assay (EMSA)

270 The CDS of *MePHD1.2* was cloned into *pMAL-c5X* for in-frame fusion with the
271 maltose binding protein (MBP). The construct of recombinant *MePHD1.2-MBP* was

272 transformed into *Escherichia coli* BL21(DE3). Induction of the corresponding protein
273 expression was completed in 100 mL luria-bertani (LB) liquid culture medium
274 containing 0.6 mmol/L isopropyl β-D-1-thiogalactopyranoside (IPTG). The cells were
275 collected after 12 h induction at 16°C. The process involved the cryogenic ultrasonic
276 lysis of bacteria, followed by protein purification via gel filtration chromatography.
277 Probe synthesis was conducted using the universal primers M13F/R with biotin
278 during PCR, and biotin labelling was applied at the 5′ end by Sangon Biotech
279 company (Shanghai, China). The same but unlabelled DNA fragment was used as a
280 competitor. EMSA was performed using a Light Shift Chemiluminescent EMSA kit
281 (Beyotime, China) and in accordance with the instructions.
282
283 Creation of genetically modified cassava
284 Gene-specific primers were used to amplify the full-length CDS of *MePHD1.2*, which
285 was subsequently fused with the modified *pCAMBIA1300* plasmid driven by the 35S
286 promoter for OE. The CRISPR/Cas9 system for gene editing was carried by the
287 *pCAMBIA1301* vector. The online website of CRISPR-P v2.0 was used to design
288 target primers that can be accomplished using the
289 (http://crispr.hzau.edu.cn/CRISPR2/). This target can simultaneously target three
290 tandem *MePHDs* in cassava. FEC was applied in a cassava genetic transformation
291 experiment involving mediation by *Agrobacterium* LBA4404. FEC cultivated on
292 Gresshof and Doy Basal Medium (GD) solid medium was inoculated into a fresh one
293 every 20 days. After infection, FEC was transfer to the nylon filter of the GD medium
294 containing carbenicillin (250 mg/L) for 3 weeks. The nylon filter with FECs was
295 transferred to the maturation and germination (MSN) medium containing 20 mg/L
296 hygromycin for 6–8 weeks under a 16 h/8 h photoperiod at 28°C (the hygromycin
297 content gradually increased from 15 mg/L to 20 mg/L). The medium was refreshed
298 every 2 weeks until the green cotyledons showed maturity. The mature cotyledons
299 were transferred to the shoot-inducting medium cassava embryo maturation (CEM)
300 (which contained 100 mg/L carbenicillin and 10 mg/L hygromycin), which was
301 refreshed every 2 weeks until complete growth of matured sprouts. Afterward, the
302 matured sprouts were cut and transferred to the MS medium.
303
304 Measurement of endogenous CGs
305 The CGs content was determined using liquid chromatography–tandem mass

306 spectrometry (LC-MS; Schmidt *et al.*, 2018).

307 Preparation of test solution

308 Standard stock solution configuration: A total of 10 mg standard substance for
309 linamarin was weighed accurately, dissolved in methanol and diluted to in a 10 mL
310 volumetric flask to prepare a 1 mg/mL standard stock solution of linamarin. A total of
311 2.5 mg standard substance for lotaustralin was weighed accurately, dissolved in
312 methanol and diluted in a 5 mL volumetric flask to prepare a 0.5 mg/mL
313 lotaustralin-labelled reserve solution.

314 Standard liquid configuration: Exactly, 100 μL 1 mg/mL linamarin stock solution and
315 200 μL 0.5 mg/mL lotaustralin stock solution were separately transferred to 10 mL
316 volumetric flasks. These stock solutions were diluted with methanol to obtain a 10
317 μg/mL mixed standard intermediate solution. A total of 1 mL mixed intermediate
318 liquid was transferred io a 10 mL volumetric flask and diluted with methanol to obtain
319 a mixed solution of 1000 ng/mL, which was used immediately after preparation.

320 Standard curve configuration: Specifically, 50, 100, 200, 500, 1000, and 5000 μL of
321 the 1000 ng/mL mixed solution were transferred to separate 10 mL volumetric flasks.
322 Each solution was diluted to a certain mark with deionised water and mixed
323 thoroughly. The mixed standard curves of linamarin and lotaustralin were prepared at
324 mass concentrations of 5, 10, 20, 50, 100, 200, and 500 ng/mL.

325 Chromatographic conditions

326 The chromatographic column used was a Shim-pack GIST C18-AQ HP (100 mm×2.1
327 mm, 1.9 μm). Mobile phase A comprised acetonitrile, and mobile phase B consisted
328 of 0.1% formic acid aqueous solution. Table 1 shows the gradient elution procedure.
329 The flow rate was 0.3 mL/min, the column temperature of the chromatographic
330 column was 0.3℃, and the injection volume was 10 μL.

331 Table 1 Gradient elution program of mobile phase

Time（min）	A（%）	B（%）
0.5	98	2
3	93	7
5	40	60
6	40	60
7	98	2

332 MS conditions

333 An AJS ESI was used as the ion source with positive scanning. Multiple reaction
334 monitoring (MRM) was utilised for CGs detection, with a capillary voltage set at
335 4000 V, nebuliser pressure at 35 psi and an ion source temperature of 350°C. Nitrogen
336 served as the sheath gas (11.5 L/min) and collision gas (1.65 L/min). Table 2 shows
337 the MRM parameters for the monitoring of each component.

338 Table 2 MRM parameter

Chemical Compound	Precursor Ion（m/z）	Fragment Ion（m/z）	Declustering Potential（V）	Collision Energy（V）
Linamarin	270	243.2	90	20
		185.1	90	16
Lotaustralin	284.1	257.1	80	20
		185.1	80	20

339 Standard curve establishment

340 Reference solutions of linamarin and lotaustralin were prepared at varying
341 concentrations. Standard curves were constructed, with the peak area and
342 concentration plotted on the vertical and horizontal axes, respectively. Linear
343 regression equations were derived for linamarin and lotaustralin. The tissue samples,
344 which were ground with liquid nitrogen, were collected from three cassava plants
345 exhibiting gene OE and *cas*, and each measurement was repeated thrice. LC-MS/MS
346 was completed by Hainan Provincial Institute of Food and Drug Inspection.
347 Measurement results were calculated using the following formula: Final
348 concentration×100 mL/Fresh weight (g)×1000.

349

350 Statistical analyses

351 Each experiment involved at least three independent replicates. The figures represent
352 the mean ± standard error (SE) of at least three independent replicates. Statistical
353 analysis was conducted using the SPSS v.21.0 software, and significance differences
354 among the treatments were assessed using Tukey's test at $P < 0.05$ or $P < 0.01$.

355

356 **Results**

357 **Isolation and bioinformatic analysis of MePHD1.2**

358 To explore the regulatory mechanism of CGs biosynthesis in cassava, we cloned and
359 used the 1505 bp promoter of *MeCYP79D2* upstream of its translation initiation site to

360 construct a bait vector. Firstly, the *pAbAi-MeCYP79D2pro* construct was linearised
361 and transformed into competent cells of Y1HGold yeast, and growth vigour of the
362 recombinant strain with *pAbAi-MeCYP79D2pro* was effectively inhibited on the
363 SD/-Ura medium containing 250 ng/ml AbA (Fig. 2a). Secondly, a mixed cDNA
364 library of storage roots, stems and leaves was used as prey to perform the Y1H assay,
365 and 1103 positive clones were screened out. A total of 109 protein-coding genes,
366 including 32 transcription factors, were identified after sequencing and functional
367 annotation (Table S3 and Fig. 2b). Among the candidates, the PHD member
368 (Manes.12G071000) appeared most frequently and accounted for approximately 5%
369 of the total positive clones.
370
371 The cassava genome contains 11 members of PHD family (Fig. S2a), with two of the
372 members arranged sequentially with Manes.12G071000 on chromosome 12. The AA
373 sequences of two other members exhibit high homology and differ only in terms of
374 specific AA residues at the C-terminal (Fig. S2b). These members were named
375 MePHD1.1 (Manes.12G070900), MePHD1.2 (Manes.12G071000) and MePHD1.3
376 (Manes.12G071100) based on their order on Chromosome 12.
377
378 PHD member is a zinc finger protein widely present in eukaryotes and belongs to the
379 zinc finger domain family; this protein performs a crucial function in transcription and
380 chromatin state regulation. MePHD1.2 presents a relatively high homology with
381 chromatin remodelling proteins, such as OsPHD, ZmPHD, HbPHD, MtPHD, etc. and
382 the highest homology with AtEBS (At4G22140) in *Arabidopsis* (Fig. 1c). *MePHD1.2*
383 encodes 216 AA residues, and its protein comprises one BAH domain and one PHD
384 domain with two Zn-binding sites (Fig. S2c). *MePHD1.2* is also expressed in all
385 tissues of cassava, with relatively high expression levels observed in mature leaves,
386 axillary buds and roots (Fig. 2d).
387
388 **Characterisation of transcription factor MePHD1.2**
389 *pCAMBIA1300-MePHD1.2-GFP* was transformed into cassava protoplasts for
390 transient expression, and an intense fluorescence of the GFP, which indicates
391 localisation of MePHD1.2 in the nucleus, was observed in the cell nucleus through
392 confocal laser microscopy (Fig. 3a). In addition, MePHD1.2 was located in the
393 nucleus of tobacco leaves (Fig. S3a). Thus, MePHD1.2 is a nuclear localised protein.

To determine whether MePHD1.2 possesses transcriptional activation capability, we coated *pGBKT7-MePHD1.2* yeast strains on SD/-Trp, SD/-Trp-His and SD/-His-Ade media. However, only the positive recombinant yeast strain with *pGBKT7-53* + *pGADm* grew well on the SD/-Trp-His and SD/-His-Ade medium (Fig. 3b). Meanwhile, the yeast strains with six truncated *MePHD1.2* segments on *pGBKT7* failed to grow on SD/-Trp-His-Ade (Figs. S3b–3c). The finding proves that MePHD1.2 lacks a transcriptional activation activity.

The *LUC* reporter gene was fused with the *5×GAL4* DNA-binding element + TATA box, with *35S::REN* as the internal reference and *GAL4BD* and *GAL4BD-PHD1.2* as effectors to perform a dual-luciferase reporter assay (Fig. 3c). *GAL4BD* exhibited a LUC/REN value 5.4 times that of *GAL4BD-PHD1.2*, which indicates that MePHD1.2 considerably reduced the expression of *LUC* reporter gene (Fig. 3d). This finding proves the transcriptional inhibitory activity of MePHD1.2.

Numerous transcription factors perform their functions through the formation of homodimers or heterodimers. The *pGBKT7-MePHD1.2* and *pGADm* constructs were cotransformed into Y2HGold yeast competent cells. Incremental concentrations of AbA were added to the yeast culture medium, and self-activation of MePHD1.2 was inhibited when the concentration of AbA reached 100 ng/mL (Fig. 3e). Next, oligomerisation analysis of MePHD1.2 was conducted, and the results show that MePHD1.2 cannot form homodimers nor heterodimers, which indicates its independent mode of function (Fig. 3f).

MePHD1.2 directly binds to the PD2 segment and represses the transcriptional activity of *MeCYP79D2*

In the examination of whether MePHD1.2 binds to the *MeCYP79D2* promoter, the yeast cells cotransformed with *pAbAi-MeCYP79D2pro* and *pGADT7-MePHD1.2* grew in the SD/-Ura-Leu medium supplemented with 300 ng/mL AbA (Fig. 4a), which suggests that MePHD1.2 can recognise and bind to the *MeCYP79D2* promoter. The 1505 bp promoter of *MeCYP79D2* was truncated into three segments, with each measuring approximately 500 bp, and named PD1, PD2 and PD3 (Fig. 4b). PD1/2/3 were used to construct bait vectors for the Y1H assay with *MePHD1.2* as prey. Firstly, only the yeast strain that contain *pAbAi-PD2* and *pGADT7-MePHD1.2* can grow on

428 the medium SD/-Ura-Leu supplemented with 50 ng/mL AbA, which implies the
429 possible binding of MePHD1.2 to the PD2 segment of the *MeCYP79D2* promoter (Fig.
430 4c).
431
432 Secondly, PD2 was labelled with 5′-biotin for the EMSA. The shift in band was
433 detected when PD2-biotin was incubated with MePHD1.2-MBP fusion protein.
434 However, no band shift occurred with MBP protein alone. The density of band shifts
435 gradually decreased with the increase in the concentration of the unlabelled
436 competitor PD2 (Fig. 4d).
437
438 Thirdly, to further clarify the regulatory effect of MePHD1.2 on the transcription of
439 *MeCYP79D2*, we used dual-luciferase assay to determine the relative luciferase
440 activity in tobacco leaves. The three negative controls had LUC/REN values 2.7 to 4.7
441 times higher than those of the cotransformation with MePHD1.2 effector and
442 *MeCYP79D2pro* reporter (Fig. 4e). Thus, MePHD1.2 can bind to the PD2 segment of
443 *MeCYP79D2* promoter and negatively regulate its transcription.
444
445 **MePHD1.2 represses the expression of *MeCYP79D2* while increases the CGs**
446 **content in *mephd1.2* mutant leaves**
447 To assess the physiological function of MePHD1.2, we developed 40 OE lines and
448 verified them by rooting test (Fig. S4a). Lines OE1, OE2 and OE3, which had
449 one-copy T-DNA according to Southern blotting and genome re-sequencing, were
450 selected (Figs. 5a–5b). These three OE lines had relatively high expressions of
451 *MePHD1.2* compared with the other OE lines, as shown by qRT-PCR assay (Fig. S4b).
452 Moreover, Hi-TOM assay identified 36 gene-edited lines as homozygous mutant, and
453 three (*cas1*, *cas2* and *cas3*) were selected because of the presence of their monoallelic
454 homozygous mutant at the target site (Figs. 5c and S4c).
455
456 To clarify the regulatory effect of MePHD1.2 on CG biosynthesis pathway, we
457 determined the contents of linamarin and lotaustralin in the first unfolded leaves of
458 OE1/2/3, *cas1/2/3* lines and CK via LC-MS. MS revealed two distinct peaks. The first
459 compound was linamarin, followed by lotaustralin (Fig. S5). The standard curve
460 equation of linamarin is $y=28.262x+2143.796$, $R^2=0.997$; that of lotaustralin is
461 $y=40.020x+1569.524$, $R^2=0.989$ (Fig. S6). The average contents of linamarin and

462 lotaustralin in the *cas* lines reached 454.02±18.60 and 85.20±23.85 mg/kg,
463 respectively, which are significantly higher than those in CK (Figs. 6a–6b). However,
464 the contents of these CGs showed no significant difference between OE lines and CK.
465 In addition, the expression levels of *MeCYP79D2* in the *cas* lines were significantly
466 higher than that in CK, but the differences between the OE lines and CK were not
467 significant (Fig. 6c). Significant differences were observed in the gene
468 expression levels between OE and *cas* lines in terms of the CG metabolism
469 pathway. The CG biosynthesis genes (*MeCYP79D2*, *MeCYP71E7* and
470 *MeUDT85K4/5*), CG decomposition genes (*MeBGLU12A/B* and *MeHNL)* and
471 CGs reutilisation genes (*MeCAS1a/b* and *MeNIT4A/B)* were all up-regulated in
472 the *cas* lines (Figs. 6d–6f). Thus, the functional deficiency of MePHD1.2 in *cas*
473 lines weakened its transcription inhibition of *MeCYP79D2*, which led to an
474 up-regulation of various CG genes and thus improved the biosynthesis efficiency of
475 CGs in the leaves.
476
477 **Discussion**
478 **MePHD1.2 can directly bind to the *MeCYP79D2* promoter and negatively**
479 **regulate its transcriptional activity**
480 PHD is a homologous domain protein with zinc fingers, and it can regulate the
481 transcriptional activity of its targets by binding to the G-rich (GNGGTG) element in
482 their promoters. Soybean GmPHDs bind to the GTGGAG element and improve stress
483 resistance capability (Winicov *et al*., 1999; Wei *et al*., 2009 and 2015). Banana
484 MaPHD1 can also directly bind to the G-rich element in *MaXTH6* promoter, which is
485 a gene related to cell-wall degradation, and down-regulate its transcriptional activity
486 to delay fruit ripening(Wei *et al*., 2017). Consistently, in this study, MePHD1.2 can
487 directly bind to the PD2 segment of *MeCYP79D2* promoter, a G-rich element located
488 in the PD2 segment, and negatively regulate its transcriptional activity (Figs. 4d–4e).
489
490 PHD proteins also act as readers to recognise histone modifications; they can bind to
491 the 'tails' of histone H3 with various modifications to recognise the histone code,
492 which results in changes in histone recognition preference to inhibit or activate DNA
493 transcription (Narro-Diego *et al*., 2017; Miura *et al*., 2020). At low temperatures,
494 vernalisation induces a state transition in flowering-related genes, such as the floral
495 integrators *FLOWERING LOCUS T* (*FT*) and *FLOWERING LOCUS C* (*FLC*), via

496 PHD-mediated chromatin remodelling, which involves histone dimerisation and
497 trimethylation. Alterations in histone methylation levels can either promote or inhibit
498 the flowering of *Arabidopsis* (Greb *et al*., 2007; Lopez-Gonzalez *et al*., 2014). The
499 PHD-type protein AL transforms the chromatin state of seed development genes,
500 namely, *ABSCISIC ACID INSENSITIVE 3* (*ABI3*) and *DELAY OF GENATION 1*
501 (*DOG1*), through the formation of a PHD-polycomb reactive complex 1 (PRC1)
502 complex with PRC1. Compression of chromatin causes inhibition in the transcription
503 activities of *ABI3* and *DOG1*, which promotes seed germination (Molitor *et al*., 2014).
504 The RNA silencing effect of the gene *ARGONAUTE 5* (*AGO5*) is inhibited by the
505 BAH-PHD-CPL2 complex (BPC), which results from the binding of BAH-PHD to
506 CPL2, a plant-specific Pol II carboxyl terminal domain phosphatase. This complex
507 enables CPL2-mediated CTD dephosphorylation to inhibit the release of Pol II from
508 the transcription start site, which results in a suppressed gene transcription (Zhang *et*
509 *al*., 2020). Further research is required to ascertain whether this complex can
510 recognise histones or recruit coregulatory factors to cooperatively regulate the
511 transcription of its targets.
512
513 **MePHD1.2 regulates CG synthesis by inhibiting the transcriptional activity of**
514 ***MeCYP79D2***
515 The variation in key AA sites of the transcription factor basic helix-loop-helix (bHLH)
516 in *Lotus Japonicus* and almond prevents them from forming dimers or affects their
517 direct recognition of promoters of downstream genes involved in CG synthesis. This
518 condition leads to a weakened or lost function, reduced transcriptional activities of
519 genes encoding the rate-limiting enzymes involved in CG biosynthesis and a
520 consequent decrease in the CG content (Sánchez-Pérez, 2019; Chen *et al*., 2022).
521 Studies on cassava have shown that the down-regulation of *MebHLH72* and
522 *MebHLH114* can lead to a decrease in the linamarin content of cassava leaves (An *et*
523 *al*., 2022). In addition, OE of bHLH transcription factors *TSAR1* or *TSAR2* in *Alfalfa,*
524 causes the increase in the expression levels of genes involved in the biosynthesis of
525 glycosides /triterpenoid saponin and rapidly increases the efficiency of their synthesis
526 (Mertens, *et al*., 2016). In this study, significantly increased expression of
527 *MeCYP79D2* and CG contents in leaves were observed in the *cas* lines of MePHD1.2
528 compared with those in CK. However, the difference between the OE lines and CK
529 was not significant. Based on these findings, knockout of *MePHD1.2* probably

530 reduced the degree of methylation in the *MeCYP79D2* promoter and prevented the
531 recruited coregulators from binding to the promoter, which inhibited their capability
532 to suppress transcriptional activity.
533
534 **CG accumulation may grant cassava the ability to tolerate barren conditions**
535 Cassava exhibits resilience to barren conditions and can achieve a certain storage root
536 yield in soils where other crops experience failure in completing their life cycle
537 (Nguyen *et al.*, 2001). Plants biosynthesise various secondary metabolites through
538 primary metabolites to perform various physiological functions; when plants suffer
539 from nutritional stress or are in a critical developmental cycle, these secondary
540 metabolites can be decomposed and assimilated back as primary metabolites to deal
541 with the new condition (Sugiyama *et al.*, 2021). CGs are nitrogen-containing
542 secondary metabolites that can be decomposed into cyanohydrins and HCN under the
543 continuous action of linamarase and hydroxynitrile hydrolase (HNL). HCN and
544 cysteine are used to synthesise cyanalanine, which is a compound catalysed by
545 cyanalanine synthase (CAS). Finally, cyanalanine is decomposed into Asp and NH_3
546 through catalysis by nitrile hydrolase (NIT4) (Picmanova *et al.*, 2015; Nielsen *et al.*,
547 2016). CGs play an important role in the resistance to herbivorous and insect
548 infestations and act as a form of nitrogen reserve for the increased content of reduced
549 nitrogen (Gladow and Woodrow, 2000). RNAi of *CYP79D1/D2* gene considerably
550 reduces cyanoside content in cassava; RNAi plants cannot complete their life cycle in
551 an environment that lacks reduced nitrogen (Siritunga and Sayre, 2003). The OE of
552 *HNL* increased the contents of Asp and other AAs in cassava roots (Siritunga *et al.*,
553 2004; Leyva-Guerrero *et al.*, 2012). Therefore, CG decomposition can increase the
554 supply of AAs and reduced nitrogen (Narayanan *et al.*, 2011; Zidenga *et al.*, 2017).
555 CGs are mainly synthesised in the leaves and transported to storage roots in cassava.
556 Nitrogen deficiency stimulates the formation and development of lateral roots to
557 increase nutrient supply (Zhang and Forde, 2000). It was speculated that the
558 transport of CGs from leaves to the roots will further strengthened, and
559 decompose into AAs and the reduced nitrogen to supply root development
560 under low nitrogen stress. Altogether, the increase in CG content of the leaves of
561 *mephd1.2* mutant maybe improve its ability to tolerate nitrogen deficiency.
562
563 **New cassava varieties with high CG content in leaves and low CG in storage**

roots can be bred with the assistance of gene editing and molecular markers

Phloem girdling test revealed the accumulation of CGs in the upper part of the girdling incision and their decreased levels in the lower part, which indicates that CGs in cassava roots are primarily transported from the leaves (Kannangara *et al*., 2011; Gleadow *et al*., 2014; Jørgensen *et al*., 2023). A weakened transport capacity can considerably reduce the content of CGs in cassava storage roots (Lieberman *et al*., 2024). Genome-wide association analysis identified two major loci influencing CG content in cassava storage roots. The multidrug and toxic compound extrusion (*MATE*) gene on chromosome 16 contributed 30% of the CG content variation in experimental populations. The G→A mutation in the fourth exon of *MATE* gene changes the original encoded Ala to Thr, which disrupts the transmembrane structure of MATE protein and substantially reduces the transport efficiency of CGs (Schmidt *et al.*, 2018; Ogbonna *et al*., 2021). The high CG content in leaves of *MePHD1.2* mutant and inefficient CG transport characteristics of parent lines can be utilised in cross-breeding and provide assistance in the molecular identification of genotypes for the rapid development of new cassava varieties with high CG content in leaves and low CGs in storage roots. Furthermore, the inclusion of these cassava varieties in agricultural production will not only improve the pest resistance of the plant and reduce the application of chemical pesticides, but also enable the production of edible cassava with low CG content in storage roots. Such achievement will be helpful in the green production of cassava and ensure food safety for populations in Asia, Africa, and Latin America.

Acknowledgements

This work was supported by grants from the China Agriculture Research System (CARS-11) and the Chinese Academy of Tropical Agricultural Sciences for the Science and Technology Innovation Team of the National Tropical Agricultural Science Center (CATASCXTD202301). Additional support was provided by the Hainan Province Graduate Innovation Research Project (Hyb2020-09) and the Key Laboratory of Tropical Fruits and Vegetables Quality and Safety for State Market Regulation (KF-2023016). We express our gratitude to Wei Wei and Shengyan Li for providing the *LUC* reporter gene with the *5×GAL4* DNA-binding element + TATA box and the *GAL4BD* vector.

Competing interests

None declared.

Author contributions

ML, YL, and XC conceptualized and designed the experiments. ML created the CRISPR-Cas9 and overexpression lines, while ML and YJL collaborated on the manuscript design and writing. ML and XZ conducted the experiments, with XYZ, QL and QY focusing on CGs content analysis. WM assisted with data analysis. WW, JL and XC originated the project idea and provided supervision throughout. All authors have reviewed and approved the final version of the manuscript.

ORCID

Mengtao Li ID https://orcid.org/0009-0007-0086-2339
Xiao Zhao ID https://orcid.org/0009-0009-2194-6384
Yuanchao Li ID https://orcid.org/0000-0001-7981-7766
Xiaoye Zhao ID https://orcid.org/0009-0008-0473-4745
Weitao Mai ID https://orcid.org/0009-0004-8356-8554
Yajun Li ID https://orcid.org/0009-0008-1634-2642
Qibing Liang ID https://orcid.org/0009-0002-0184-0135
Qingchun Yin ID https://orcid.org/0000-0003-4958-5663
Wenquan Wang ID https://orcid.org/0000-0002-7603-3437
Jinping Liu ID https://orcid.org/0000-0001-5388-6026
Xin Chen ID https://orcid.org/0000-0001-8189-9967

Data availability

All data supporting the findings are included in this manuscript. Additional supporting informations can be found in the Supporting Information section at the end of the article.

References

An F, X XH, Chen T, Xue JJ, Luo XQ, Ou WJ, Li KM, Cai J, Chen SB. 2022. Systematic analysis of bHLH transcription factors in cassava uncovers their roles in postharvest physiological deterioration and cyanogenic glycosides biosynthesis. *Frontiers in Plant*

632 *Science* **13**: 901128.
633 **Andersen MD, Busk PK, Svendsen I, Møller BL. 2000.** Cytochromes P-450 from cassava
634 (*Manihot esculenta* Crantz) catalyzing the first steps in the biosynthesis of the cyanogenic
635 glucosides linamarin and lotaustralin: cloning, functional expression in pichia pastoris, and
636 substrate specificity of the isolated recombinant enzymes. *Journal of Biological Chemistry*
637 **275**:1966-1975.
638 **Bak S, Kahn RA, Nielsen HL, Møller BL, Halkier BA. 1998.** Cloning of three A-type
639 cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) Moench by a
640 PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a
641 multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin.
642 *Plant molecular biology* **36**: 393-405.
643 **Ballhorn DJ, Kautz S, Laumann JM. 2016.** Herbivore damage induces a transgenerational
644 increase of cyanogenesis in wild lima bean (*Phaseolus lunatus*). *Chemoecology* **26**:1-5.
645 **Banea-Mayambu JP, Tylleskär T, Gitebo N, Matadi N, Gebre-Medhin M, Rosling H. 1997.**
646 Geographical and seasonal association between linamarin and cyanide exposure from cassava
647 and the upper motor neurone disease konzo in former Zaire. *Tropical Medicine &*
648 *International Health* **2**:1143-1151.
649 **Chen C, Liu F, Zhang K, Niu X, Zhao H, Liu Q, Georgiev MI, Xu X, Zhang X, Zhou M.**
650 **2022.** MeJA-responsive bHLH transcription factor LjbHLH7 regulates cyanogenic glucoside
651 biosynthesis in *Lotus japonicus*. *Journal of experimental botany* **18**: 2650-2665.
652 **Dudareva N, Pichersky E, Gershenzon J. 2004.** Biochemistry of plant volatiles. *Plant*
653 *physiology* **135**:1893-902..
654 **Easson ML, Malka O, Paetz C, Hojná A, Reichelt M, Stein B, van Brunschot S, Feldmesser**
655 **E, Campbell L, Colvin J, Winter S. 2021.** Activation and detoxification of cassava
656 cyanogenic glucosides by the whitefly *Bemisia tabaci*. *Scientific reports* **24**: 13244.
657 **Fischbach MA, Clardy J. 2007.** One pathway, many products. *Nature chemical biology*
658 **3**:353-355.
659 **Forslund K, Morant M, Jørgensen B, Olsen CE, Asamizu E, Sato S, Tabata S, Bak S. 2004.**
660 Biosynthesis of the nitrile glucosides rhodiocyanoside A and D and the cyanogenic
661 glucosides lotaustralin and linamarin in *Lotus japonicus*. *Plant physiology* **135**: 71-84.
662 **Ganjewala D. 2010.** Advances in cyanogenic glycosides biosynthesis and analyses in plants: A
663 review. *Acta Biologica Szegediensis* **54**: 1-4.
664 **Gleadow RM, Møller BL. 2014.** Cyanogenic glycosides: synthesis, physiology, and phenotypic
665 plasticity. *Annual review of plant biology* **65**: 155-85.
666 **Gleadow RM, Woodrow IE. 2000.** Temporal and spatial variation in cyanogenic glycosides in
667 Eucalyptus cladocalyx. *Tree Physiology* 20: 591-598.
668 **Gomez MA, Berkoff KC, Gill BK, Iavarone AT, Lieberman SE, Ma JM, Schultink A,**
669 **Karavolias NG, Chauhan RD, Taylor NJ, Staskawicz BJ. 2023.** CRISPR-Cas9-mediated

670 knockout of *CYP79D1* and *CYP79D2* in cassava attenuates toxic cyanogen production.
671 *Frontiers in Plant Science* **13**: 1079254.
672 **Greb T, Mylne JS, Crevillen P, Geraldo N, An H, Gendall AR, Dean C. 2007.** The PHD finger
673 protein VRN5 functions in the epigenetic silencing of Arabidopsis FLC. *Current biology* **17**:
674 73-78.
675 **Jain K, Fraser CS, Marunde MR, Parker MM, Sagum C, Burg JM, Hall N, Popova IK,**
676 **Rodriguez KL, Vaidya A, Krajewski K. 2020.** Characterization of the plant homeodomain
677 (PHD) reader family for their histone tail interactions. *Epigenetics & chromatin* **13**: 1-11.
678 **Jørgensen K, Bak S, Busk PK, Sørensen C, Olsen CE, Puonti-Kaerlas J, Møller BL. 2005.**
679 Cassava plants with a depleted cyanogenic glucoside content in leaves and tubers.
680 Distribution of cyanogenic glucosides, their site of synthesis and transport, and blockage of
681 the biosynthesis by RNA interference technology. *Plant Physiology* **139**: 363-374.
682 **Jørgensen K, Morant AV, Morant M, Jensen NB, Olsen CE, Kannangara R, Motawia MS,**
683 **Møller BL, Bak S. 2011.** Biosynthesis of the cyanogenic glucosides linamarin and
684 lotaustralin in cassava: isolation, biochemical characterization, and expression pattern of
685 CYP71E7, the oxime-metabolizing cytochrome P450 enzyme. *Plant physiology* **155**:
686 282-292.
687 **Kannangara R, Motawia MS, Hansen NK, Paquette SM, Olsen CE, Møller BL, Jørgensen K.**
688 **2011.** Characterization and expression profile of two UDP-glucosyltransferases, UGT85K4
689 and UGT85K5, catalyzing the last step in cyanogenic glucoside biosynthesis in cassava. *The*
690 *Plant Journal* **68**: 287-301.
691 **Kongsawadworakul P, Viboonjun U, Romruensukharom P, Chantuma P, Ruderman S,**
692 **Chrestin H. 2009.** The leaf, inner bark and latex cyanide potential of Hevea brasiliensis:
693 evidence for involvement of cyanogenic glucosides in rubber yield. *Phytochemistry* **70**:
694 730-739.
695 **Lai D, Maimann AB, Macea E, Ocampo CH, Cardona G, Pičmanová M, Darbani B, Olsen**
696 **CE, Debouck D, Raatz B, Møller BL. 2020.** Biosynthesis of cyanogenic glucosides in
697 Phaseolus lunatus and the evolution of oxime-based defenses. *Plant Direct* **4**: e00244.
698 **Latif S, Müller J. 2015.** Potential of cassava leaves in human nutrition: A review. *Trends in Food*
699 *Science & Technology* **44**: 147-158.
700 **Lebot V. 2019.** Tropical root and tuber crops. *Cabi*.
701 **Leyva-Guerrero E, Narayanan NN, Ihemere U, Sayre RT. 2012.** Iron and protein
702 biofortification of cassava: lessons learned. *Current opinion in biotechnology* **23**: 257-264.
703 **Lieberman SE, Gueorguieva GA, Gill BK, Litvak L, Cruz AG, Lyons JB, Cho MJ,**
704 **Karavolias N. 2024.** Transporter editing in cassava indicates local production of cyanogenic
705 glucosides in, and export from, cassava roots. *Plant Biotechnology Journal* **22**: 790.

706 **López-González L, Mouriz A, Narro-Diego L, Bustos R, Martínez-Zapater JM, Jarillo JA,**
707 **Piñeiro M. 2014.** Chromatin-dependent repression of the Arabidopsis floral integrator genes
708 involves plant specific PHD-containing proteins. *The Plant Cell* **26**: 3922-3938.
709 **Ma PA, Chen X, Liu C, Xia ZQ, Song Y, Zeng CY, Li YZ,Wang WQ. 2018.** MePHD1 as a
710 PHD-finger protein negatively regulates ADP-glucose pyrophosphorylase small subunit1a
711 gene in cassava. *International Journal of Molecular Sciences* **19**: 2831.
712 **McMahon J, Sayre R, Zidenga T. 2022.** Cyanogenesis in cassava and its molecular manipulation
713 for crop improvement. *Journal of Experimental Botany* **73**: 1853-1867.
714 **Mertens J, Pollier J, Vanden Bossche R, Lopez-Vidriero I, Franco-Zorrilla JM, Goossens A.**
715 **2016.** The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin
716 biosynthesis in *Medicago truncatula*. *Plant physiology* **170**: 194-210.
717 **Miura K, Renhu N, Suzaki T. 2020.** The PHD finger of *Arabidopsis* SIZ1 recognizes
718 trimethylated histone H3K4 mediating SIZ1 function and abiotic stress response.
719 *Communications biolog*y **3**: 23.
720 **Molitor AM, Bu Z, Yu Y, Shen WH. 2014.** *Arabidopsis* AL PHD-PRC1 complexes promote seed
721 germination through H3K4me3-to-H3K27me3 chromatin state switch in repression of seed
722 developmental genes. *PLoS genetics* **10**: e1004091.
723 **Møller BL. 2010.** Functional diversifications of cyanogenic glucosides. *Current opinion in plant*
724 *biology* **13**: 337-346.
725 **Narayanan NN, Ihemere U, Ellery C, Sayre RT. 2011.** Overexpression of hydroxynitrile lyase
726 in cassava roots elevates protein and free amino acids while reducing residual cyanogen
727 levels. *PloS one* **6**: e21996.
728 **Narro-Diego L, López-González L, Jarillo JA, Piñeiro M. 2017.** The PHD‐containing protein
729 EARLY BOLTING IN SHORT DAYS regulates seed dormancy in A rabidopsis. *Plant, Cell*
730 *& Environment* **40**:2393-2405.
731 **Nguyen H, Schoenau JJ, Van Rees KC, Nguyen D, Qian P. 2001.** Long-term nitrogen,
732 phosphorus and potassium fertilization of cassava influences soil chemical properties in
733 North Vietnam. *Canadian journal of soil science* **81**: 481-488..
734 **Nielsen LJ, Stuart P, Pičmanová M, Rasmussen S, Olsen CE, Harholt J, Møller BL,**
735 **Bjarnholt N. 2016.** Dhurrin metabolism in the developing grain of *Sorghum bicolor* (L.)
736 Moench investigated by metabolite profiling and novel clustering analyses of time-resolved
737 transcriptomic data. *BMC genomics* **17**: 1-24.
738 **Ogbonna AC, Braatz de Andrade LR, Rabbi IY, Mueller LA, Jorge de Oliveira E, Bauchet**
739 **GJ. 2021.** Large-scale genome-wide association study, using historical data, identifies
740 conserved genetic architecture of cyanogenic glucoside content in cassava (*Manihot*
741 *esculenta* Crantz) root. *The Plant Journal* **105**: 754-770.
742 **Panghal A, Munezero C, Sharma P, Chhikara N. 2019.** Cassava toxicity, detoxification and its
743 food applications: a review. *Toxin Reviews*.

744 **Pičmanová M, Neilson EH, Motawia MS, Olsen CE, Agerbirk N, Gray CJ, Flitsch S, Meier S,**
745 **Silvestro D, Jørgensen K, Sánchez-Pérez R. 2015.** A recycling pathway for cyanogenic
746 glycosides evidenced by the comparative metabolic profiling in three cyanogenic plant
747 species. *Biochemical Journal* **469**: 375-89.
748 **Qian F, Zhao QY, Zhang TN, Li YL, Su YN, Li L, Sui JH, Chen S, He XJ. 2021.** A histone
749 H3K27me3 reader cooperates with a family of PHD finger-containing proteins to regulate
750 flowering time in *Arabidopsis*. *Journal of integrative plant biology* **63**: 787-802..
751 **Sánchez-Pérez R, Pavan S, Mazzeo R, Moldovan C, Aiese Cigliano R, Del Cueto J, Ricciardi**
752 **F, Lotti C, Ricciardi L, Dicenta F, López-Marqués RL. 2019.** Mutation of a bHLH
753 transcription factor allowed almond domestication. *Science* **364**:1095-1098.
754 **Santana MA, Vásquez V, Matehus J, Aldao RR. 2002.** Linamarase expression in cassava
755 cultivars with roots of low-and high-cyanide content. *Plant physiology* **129**: 1686-1694.
756 **Schmidt FB, Cho SK, Olsen CE, Yang SW, Møller BL, Jørgensen K. 2018.** Diurnal regulation
757 of cyanogenic glucoside biosynthesis and endogenous turnover in cassava. *Plant Direct* **2**:
758 e00038.
759 **Schmidt FB, Heskes AM, Thinagaran D, Møller BL, Jørgensen K, Boughton BA. 2018.** Mass
760 spectrometry based imaging of labile glucosides in plants. *Frontiers in Plant Science* **9**:
761 355425.
762 **Shang Y, Ma Y, Zhou Y, Zhang H, Duan L, Chen H, Zeng J, Zhou Q, Wang S, Gu W, Liu M.**
763 **2014.** Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* **346**:
764 1084-1088.
765 **Siritunga D, Arias-Garzon D, White W, Sayre RT. 2004.** Over-expression of hydroxynitrile
766 lyase in transgenic cassava roots accelerates cyanogenesis and food detoxification. *Plant*
767 *Biotechnology Journal* **2**: 37-43.
768 **Siritunga D, Sayre RT. 2003.** Generation of cyanogen-free transgenic cassava. *Planta* **217**:
769 367-73.
770 **Sugiyama R, Li R, Kuwahara A, Nakabayashi R, Sotta N, Mori T, Ito T, Ohkama-Ohtsu N,**
771 **Fujiwara T, Saito K, Nakano RT. 2021.** Retrograde sulfur flow from glucosinolates to
772 cysteine in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **118**:
773 e2017890118.
774 **Szczyglowski K, Hamburger D, Kapranov P, de Bruijn FJ. 1997.** Construction of a Lotus
775 japonicus late nodulin expressed sequence tag library and identification of novel
776 nodule-specific genes. *Plant physiology* **114**: 1335-1346.
777 **Wei W, Fan ZQ, Chen JY, Kuang JF, LU WJ, Shan W. 2017.** A banana PHD-type transcription
778 factor MaPHD1 represses a cell wall-degradation gene *MaXTH6* during fruit ripening.
779 *Horticultural plant journal* **3**:190-198.
780 **Wei W, Huang J, Hao YJ, Zou HF, Wang HW, Zhao JY, Liu XY, Zhang WK, Ma B, Zhang**
781 **JS, Chen SY. 2009.** Soybean GmPHD-type transcription regulators improve stress tolerance

782 in transgenic Arabidopsis plants. *PloS one* **4**: e7209.
783 **Wei W, Zhang YQ, Tao JJ, Chen HW, Li QT, Zhang WK, Ma B, Lin Q, Zhang JS, Chen SY.**
784 **2015.** The Alfin-like homeodomain finger protein AL5 suppresses multiple negative factors
785 to confer abiotic stress tolerance in Arabidopsis. *The Plant Journal* **81**: 871-883.
786 **Winicov I, Bastola DR. 1999.** Transgenic overexpression of the transcription factor Alfin1
787 enhances expression of the endogenous *MsPRP2* gene in alfalfa and improves salinity
788 tolerance of the plants. *Plant physiology* **120**: 473-480.
789 **Zagrobelny M, Bak S, Rasmussen AV, Jørgensen B, Naumann CM, Møller BL. 2004.**
790 Cyanogenic glucosides and plant–insect interactions. *Phytochemistry* **65**: 293-306.
791 **Zhang P, Bohl-Zenger S, Puonti-Kaerlas J, Potrykus I, Gruissem W. 2003.** Two cassava
792 promoters related to vascular expression and storage root formation. *Planta* **218**:192-203.
793 **Zhang H, Forde BG. 2000.** Regulation of Arabidopsis root development by nitrate availability.
794 *Journal of experimental botany* **1**: 51-59.
795 **Zhang YZ, Yuan JL, Zhang LR, Chen CX, Wang YH, Zhang GP, Li Peng, Xie SS, Jiang J,**
796 **Zhu JK, Du JM, Duan CG. 2020.** Coupling of H3K27me3 recognition with transcriptional
797 repression through the BAH-PHD-CPL2 complex in *Arabidopsis*. *Nature communications* **11**:
798 6212.
799 **Zidenga T, Siritunga D, Sayre RT. 2017.** Cyanogen metabolism in cassava roots: impact on
800 protein synthesis and root development. *Frontiers in Plant Science* **8**: 241486.
801

802 **Supporting Information**

803 Additional supporting information may be found online in the Supporting Information
804 section at the end of the article.
805 Fig. S1 Promoter sequence (2 kb) of *MeCYP79D2* gene.
806 Fig. S2 Characterisation of MePHD1.2 in cassava.
807 Fig. S3 Subcellular localisation and transcription activation region assay of
808 MePHD1.2.
809 Fig. S4 Identification and phenotypic characterisation of *MePHD1.2* transgenic
810 plants.
811 Fig. S5 MS peak of linamarin and lotaustralin determination.
812 Fig. S6 Standard curves of linamarin and lotaustralin in LC-MS/MS detection.
813 Table S1 Part of cis element on the *MeCYP79D2* promoter.
814 Table S2 Primers used for vector construction in this study.
815 Table S3 Partially interacting proteins identified via Y1H assay.

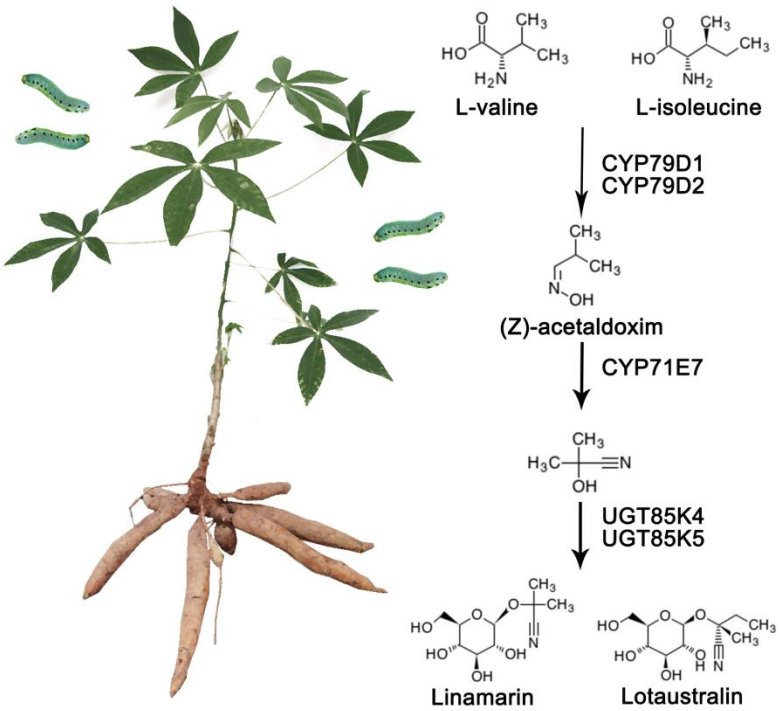

Fig. 1 CG biosynthesis pathway in cassava.

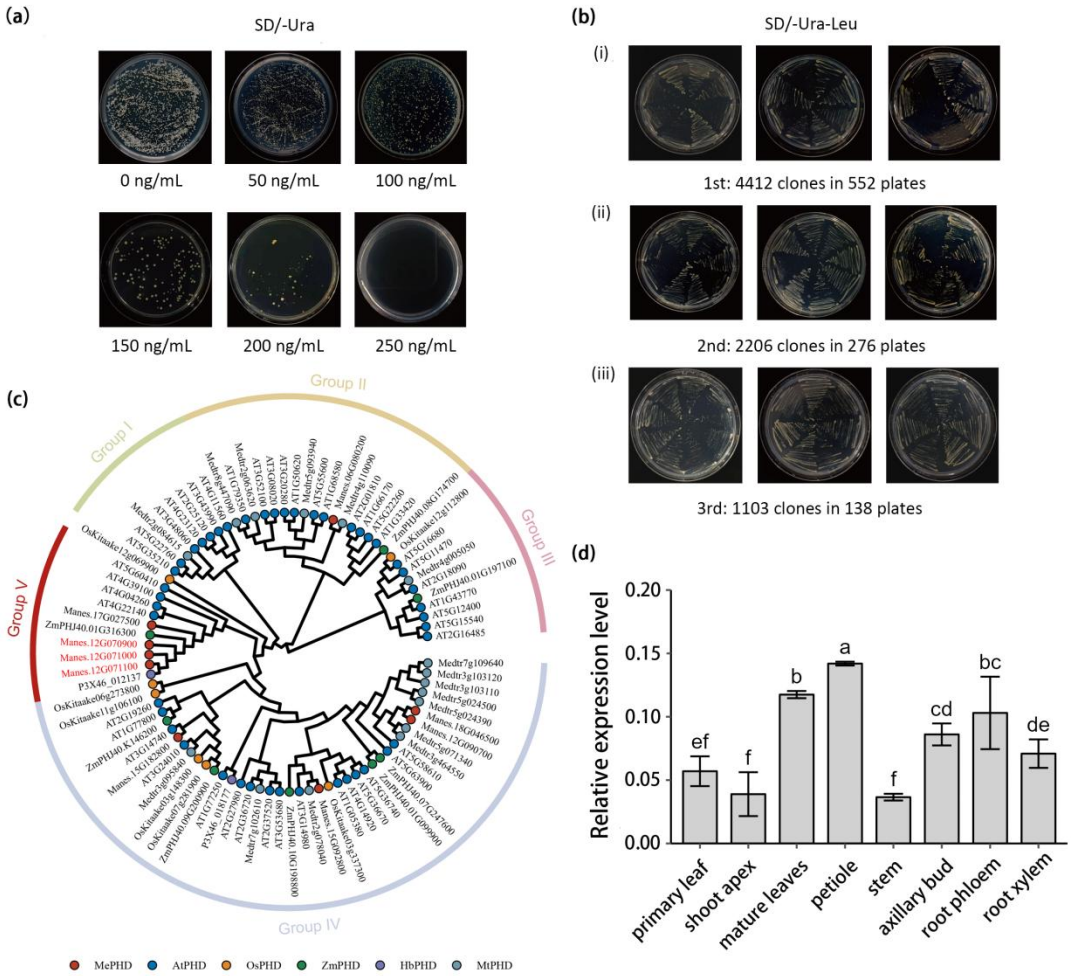

Fig. 2 Yeast one-hybrid assay and characterisation of PHD transcription factors. (a) Inhibition of the self-activation of *pAbAi-MeCYP79D2*. AbA was added to SD/-Ura medium at concentrations of 50, 100, 150, 200 and 250 ng/mL. (b) Verify 3 times to confirm the positive clones. (c) Phylogenetic analysis of homologous PHD transcription factors from various species. AA sequences were aligned using MEGA6 software, and a phylogenetic tree was constructed via the NJ method. MePHD/cassava, AtPHD/arabidopsis, OsPHD/rice, ZmPHD/maize, HbPHD/rubber tree and MtPHD/alfalfa. (d) Relative expression levels of *MePHD1.2* in various tissues of SC8 cassava variety. The data are the means ± standard deviation (SD) (n=3). Statistically significant differences were determined via one-way analysis of variance (ANOVA). Different letters indicate statistically significant differences.

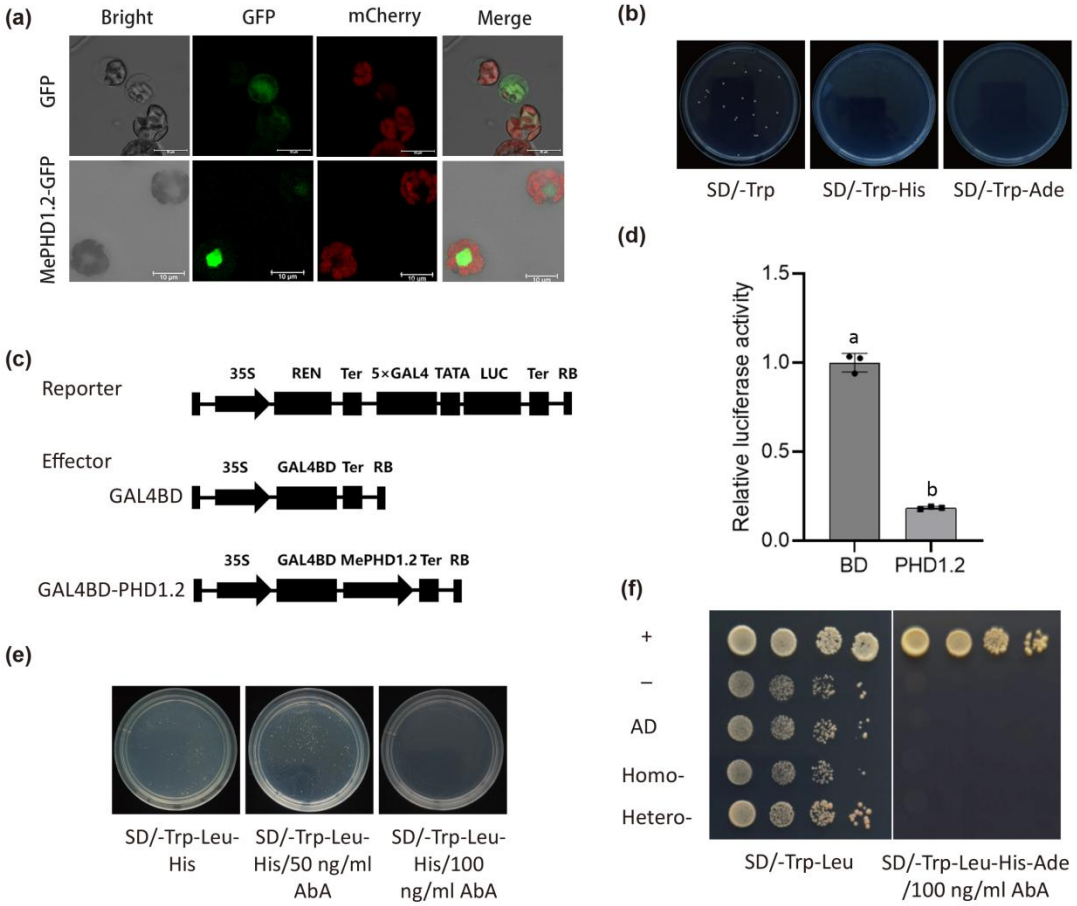

Fig. 3 MePHD1.2 is a transcription factor with a transcriptional inhibitory activity. (a) Subcellular localisation of MePHD1.2. Full-length MePHD1.2 fused with the improved GFP and the nucleus marker mCherry were coexpressed in cassava protoplast cells. *pSL1 Plus-GFP* served as a negative control. Scale bar: 10 μm. (b) Validation of MePHD1.2 transcription activation activity in yeast cells. The *pGBKT7-MePHD1.2* recombinant plasmid was transformed into Y2HGold competent cells, and the bacterial solution was diluted to OD_{600}=0.002. A 100 μL yeast liquid was obtained and subsequently applied to the SD/-Trp, SD/-Trp-His and SD/-Trp-Ade media. The plates were incubated at 30°C for 2–3 days in an incubator. (c) Transcription activation activity reporting system model. (d) Relative luciferase activity was detected in the cassava protoplast. The data are the mean ± SD (n=3). One-way ANOVA revealed statistically significant differences. Different letters denote statistically significant differences. (e) Inhibition of the self-activation of MePHD1.2 transcription factor. (f) Oligomerisation analysis of MePHD1.2 transcription factor. Cotransformation of *pGBKT7-MePHD1.2* with homologous *pGAD-MePHD1.2* and heterologous *pGAD-MePHD1.1* was conducted in yeast

Y2HGold, followed by culture on SD/-Trp-Leu-His-Ade media containing 100 ng/mL AbA. '+' represents the combination of *pGBKT7-53* and *pGADT7*, '-' denotes the combination of *pGBKT7-Lam* and *pGADT7*, 'AD' indicates the combination of *pGBKT7-MePHD1.2* and *pGADT7*, 'Homo –' refers to homologous interaction and 'Hetero –' means heterologous interaction.

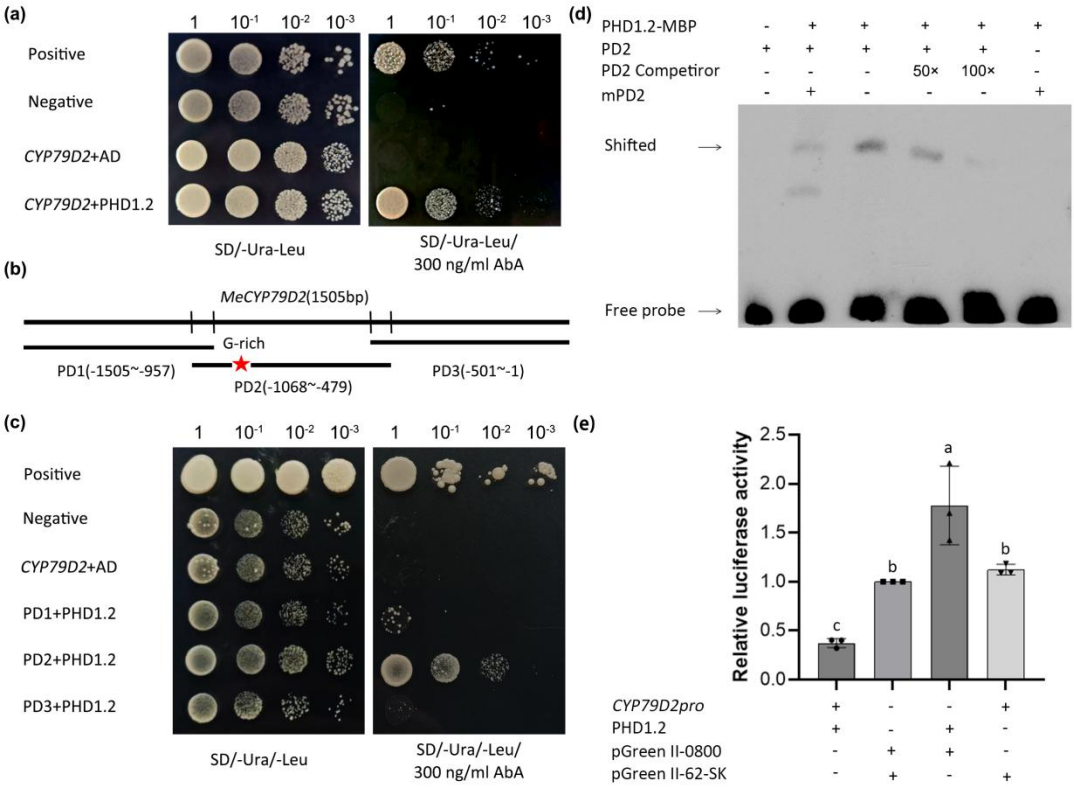

Fig. 4 Binding of MePHD1.2 to the PD2 segment of *MeCYP79D2* promoter and its repressive action on transcription activity. (a) Yeast one-hybrid assay between MePHD1.2 and *MeCYP79D2* promoter. *pGADT7-MePHD1.2* was transformed into yeast competent cells containing *pAbAi-MeCYP79D2pro*. The cells were cultured at various gradients of bacterial concentrations on SD/-Ura-Leu medium supplemented with 300 ng/mL AbA. Positive and negative controls represented the combinations of *pAbAi-p53* and *pGADT7* and that of *pAbAi* and *pGADT7*, respectively. AD indicates *pGADT7*. (b) Pattern diagram of truncated promoter segments. PD1, PD2 and PD3 represent the three segments of *MeCYP79D2*. (c) Yeast one-hybrid assay between MePHD1.2 and PD1/2/3. (d) EMSA of MePHD1.2 bound to the PD2 segment *in vitro*. Biotin-labelled DNA probes (PD2 or mPD2) were incubated with the MBP-MePHD1.2 protein, and DNA–protein complexes were separated on 6% native polyacrylamide gels. The cold probes of the competitor were the unlabelled probe;

‘50×’ and ‘100×’ represent 50- and 100-fold molar excess, respectively. The symbol ‘+’ or ‘-’ indicates the presence or absence of the corresponding proteins and probes, respectively. (e) Dual-luciferase assay. The LUC/REN ratio, which was set to 1, was normalised to the samples combined with *pGreenII 62-SK* and *pGreenII 0800-LUC*. LUC, firefly luciferase activity; REN, Renilla luciferase activity. The data are the mean ± SD (n=3). Statistically significant differences were determined using one-way ANOVA. Different letters denote statistically significant differences.

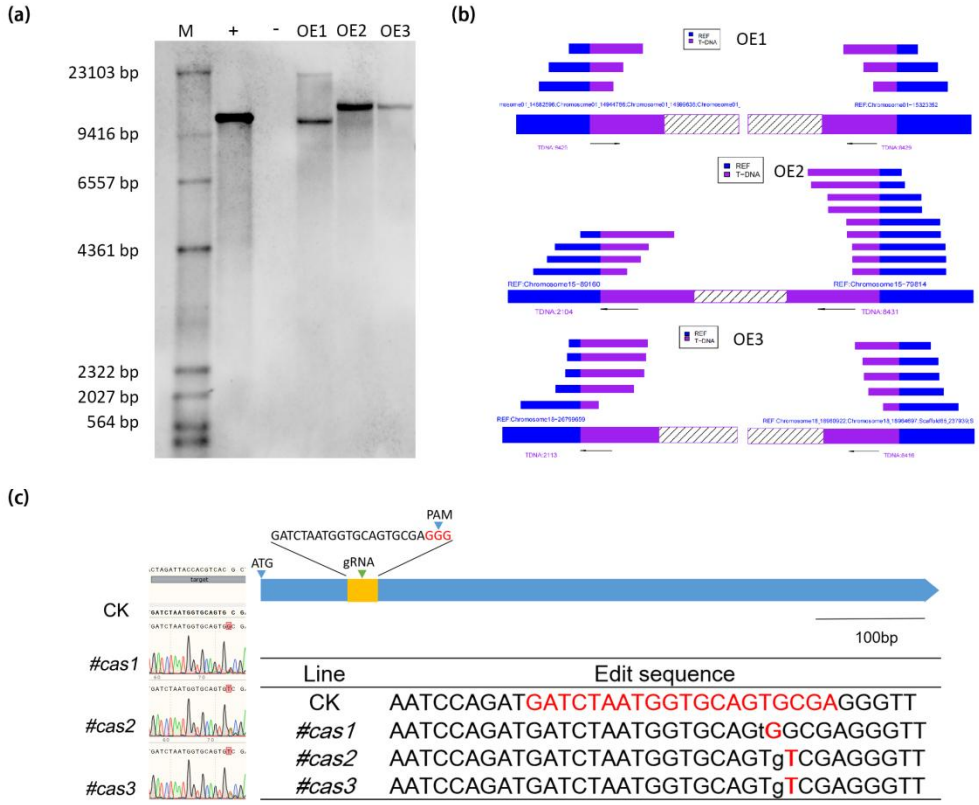

Fig. 5 Identification of OE and *cas* lines. (a) Southern blotting of *MePHD1.2* OE lines. (b) Genome resequencing alignment of *MePHD1.2* OE lines. (c) Mutant types of *MePHD1.2 cas* lines. CK represents the regenerated plant of SC8 infected with an empty vector in *Agrobacterium.*

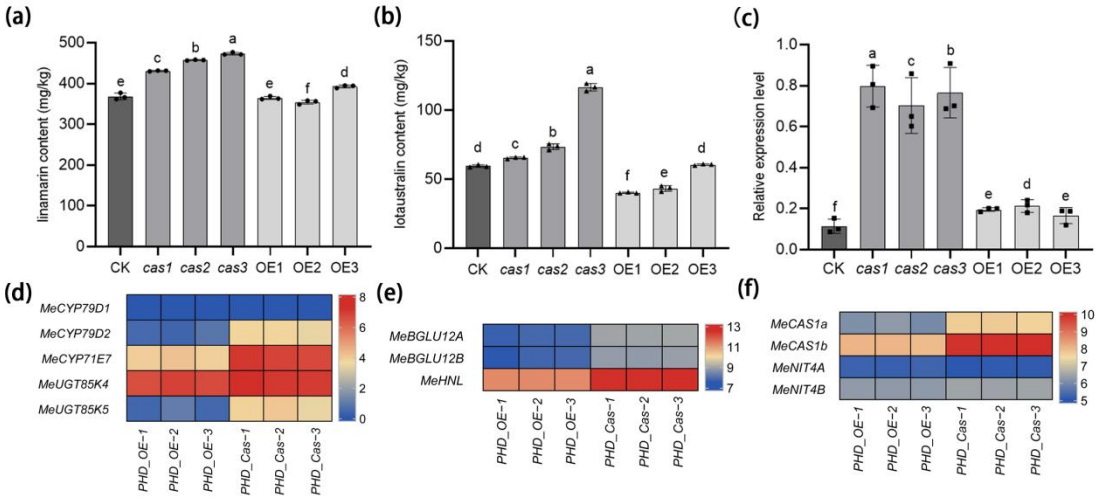

Fig. 6 Determination of CG content and relative expression levels of *MeCYP79D2*. (a–b) Linamarin and lotaustralin contents in CK and *cas*/OE lines. (c) Expression levels of *MeCYP79D2* in CK and *cas*/OE lines. (d–f) Expression differences of key genes related to CG biosynthesis, decomposition and reutilisation between OE and *cas* lines. The data are the mean ± SD (n=3). One-way ANOVA revealed statistically significant differences. Different letters indicate statistically significant differences.