Identification of basic helix-loop-helix transcription factors that activate betulinic acid biosynthesis by RNA-sequencing of hydroponically cultured *Lotus japonicus*

Hayato Suzuki[1,2], Hirokazu Takahashi[3], Ery Odette Fukushima[1,4], Mikio Nakazono[3], Toshiya Muranaka[1,5], and Hikaru Seki[1,5*]

[1] Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

[2] Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Higashi 1-1-1, Tsukuba, Ibaraki 305-8562, Japan

[3] Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa, Nagoya 464-8601, Japan

[4] Plant Translational Research Group, Universidad Regional Amazónica IKIAM, Muyuna kilómetro 7 vía a Alto Tena 150101, Ecuador

[5] Industrial Biotechnology Initiative Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

* Corresponding author (hseki@bio.eng.osaka-u.ac.jp; Tel +81-6-6879-7423)

ORCIDs

Hayato Suzuki, https://orcid.org/0000-0001-6120-9471

Hirokazu Takahashi, https://orcid.org/0000-0001-6182-3236

Ery Odette Fukushima, https://orcid.org/0000-0003-4099-4292

Mikio Nakazono, https://orcid.org/0000-0001-7119-2052

Toshiya Muranaka, https://orcid.org/0000-0003-1058-2473

Hikaru Seki, https://orcid.org/0000-0002-6426-0368

Word counts: Summary (198), Main body in total (6293), Introduction (592), Materials and Methods (2392), Results (1883), Discussion (1426)

Number of: Figures (7; Figures other than Fig. 3 will be coloured), Tables (2)

Type and number of supporting information: Figures (9); Tables (7)

Summary

- Although triterpenes are ubiquitous in plant kingdom, their biosynthetic regulatory mechanisms are limitedly understood. Here, we found that hydroponic culture dramatically activated betulinic acid (BA) biosynthesis in the model Fabaceae *Lotus japonicus*, and investigated its transcriptional regulation.
- Fabaceae plants develop secondary aerenchyma (SA) on the surface of hypocotyls and roots during flooding for root air diffusion. Hydroponic culture induced SA in *L. japonicus* and simultaneously increased the accumulation of BA and the transcript levels of its biosynthetic genes. RNA-sequencing of soil-grown and hydroponically cultured plant tissues, including SA collected by laser microdissection, revealed that several transcription factor genes were co-upregulated with BA biosynthetic genes.
- Overexpression of *LjbHLH32* and *LjbHLH50* in *L. japonicus* transgenic hairy roots upregulated the expression of BA biosynthesis genes, resulting in enhanced BA accumulation. However, transient luciferase reporter assays in *Arabidopsis* mesophyll cell protoplasts showed that LjbHLH32 transactivated promoters of biosynthetic genes in the soyasaponin pathway but not the BA pathway, like its homolog GubHLH3, a soyasaponin biosynthesis regulator in *Glycyrrhiza uralensis*. This suggested the evolutionary origin and complex regulatory mechanisms of BA biosynthesis in Fabaceae.
- This study sheds light on the unrevealed biosynthetic regulatory mechanisms of triterpenes in Fabaceae plants.

Key words: betulinic acid, bHLH, hydroponics, laser microdissection (LMD), *Lotus japonicus*, secondary aerenchyma, transcription factor, triterpene

Introduction

Triterpenes are plant specialized (secondary) metabolites with diverse structures and biological properties (Thimmappa *et al.*, 2014). The common precursor, 2,3-oxidosqualene, is composed of six isoprene units supplied by the mevalonate (MVA) pathway. Multiple oxidosqualene cyclases (OSCs), cytochrome P450 monooxygenases (P450s), and UDP-dependent glycosyltransferases (UGTs) catalyze conversion of the substrates to structurally diversified non-glycosylated triterpenes and triterpene glycosides (saponins; Seki *et al.*, 2015).

Fabaceae plants produce a large amount of triterpenes with marked chemodiversity. Soyasaponins are specific to Fabaceae plants such as *Glycine max*, *Medicago* spp., *Glycyrrhiza* spp., and *Lotus japonicus* (Fenwick, 1992; Hayashi *et al.*, 1998; Tsuno *et al.*, 2018). *Medicago truncatula* produces oleanolic acid (OA)-derived hemolytic saponins (Huhman *et al.*, 2005). Thickened roots of *Glycyrrhiza uralensis* accumulate glycyrrhizin, but more OA and betulinic acid (BA) were produced in tissue-cultured stolons (Kojoma *et al.*, 2010; Tamura *et al.*, 2017). *L. japonicus* accumulates C-28-oxidized triterpenes [ursolic acid (UA), OA, and BA] in a tissue-dependent manner (Suzuki *et al.*, 2019). Biosynthesis pathways of these Fabaceae triterpenes have been almost understood by the recent identification of CSyGT (cellulose-synthase-derived glycosyltransferase) enzymes catalyzing the glucuronosylation of triterpene aglycones (Jozwiak *et al.*, 2020; Chung *et al.*, 2020). However, their physiological functions and biosynthetic regulatory mechanisms are unclear.

Plant genomes possess a large number of basic helix-loop-helix (bHLH)-type transcription factors (TFs), which are classified into approximately 25 clades based on

sequence similarity in the bHLH domain and other conserved domains (Heim *et al.*, 2003). Fabaceae plants possess more clade IVa bHLH genes than related plant families, which are subdivided into three subclades (IVa1-3) based on a phylogenetic tree of the full-length proteins (Suzuki *et al.*, 2021). Some members of subclade IVa1 are regulators of Fabaceae triterpene saponin biosynthesis. TSAR1 and TSAR2 are positive regulators of soyasaponins and hemolytic saponins, respectively, in *M. truncatula* (Mertens *et al.*, 2016). TSAR3, their paralog, is a regulator of hemolytic saponin biosynthesis during seed development (Ribeiro *et al.*, 2020). GubHLH3, a TSAR homolog, upregulated soyasaponin biosynthesis in *G. uralensis* (Tamura *et al.*, 2018). TFs regulating BA biosynthesis have yet been identified in Fabaceae plants. On the other hand, BpbHLH9 (clade Ia) was identified as an OA and BA biosynthesis regulator in the Betulaceae plant *Betula platyphylla* (Yin *et al.*, 2017). The clade Ib bHLH TFs Bl and Bt are regulators of cucurbitane-type triterpenes in leaves and fruits, respectively, of cucumber (Cucurbitaceae; Shang *et al.*, 2014). Thus, identification of triterpene-related bHLH TFs in *L. japonicus* provides insight into the evolutionary history of triterpene biosynthesis regulators in the Fabaceae and other plant families.

We performed triterpene profiling to characterize triterpene production in *L. japonicus*. A large amount of BA accumulated in secondary aerenchyma (SA) tissue in hypocotyl and roots when cultured hydroponically. SA is a white, spongy tissue developed from phellogen [secondary meristem (SM)], whose formation is an adaptive trait of Fabaceae plants to survive flooding by promoting oxygen diffusion in roots (Mochizuki *et al.*, 2000; Yamauchi *et al.*, 2013). Because the expression of BA biosynthetic genes (*OSC3* and *CYP716A51*) was upregulated in whole roots of hydroponically cultured plants and SA, we performed comparative RNA sequencing

103 (RNA-seq) analysis to identify candidate TFs that positively regulate BA biosynthesis.
104 From comparisons between soil-cultured and hydroponically cultured whole roots and
105 between SA and SM collected by laser microdissection (LMD), we identified 23 TFs
106 co-upregulated in hydroponic roots and SA. Among them, we focused on LjbHLH32
107 (subclade IVa2) and LjbHLH50 (clade Ia) based on their co-expression with BA
108 biosynthetic genes and phylogenetic analysis. The functions of these two bHLH-type
109 TFs in triterpene biosynthesis regulation were examined and are discussed.

110 **Materials and Methods**

111 **Plant materials**

112 Seeds of *L. japonicus* Gifu B-129 accession and *M. truncatula* cv. R108 and Jemalong
113 A17 were sterilized using a five-fold-diluted sodium hypochlorite solution containing
114 0.02% Tween-20 for 15 min with mild agitation. After washing three times in sterilized
115 water, they were soaked in water with mild agitation overnight. The seeds were
116 germinated on 0.8% agar with incubation for 4 days (*L. japonicus*) or 3 days (*M.*
117 *truncatula*) in the dark and for 2 days on a 16 h light/8 h dark cycle at 25°C. The
118 seedlings were transplanted to soil pots or hydroponic solutions and cultured in a plant
119 growth room maintained at 23°C until harvest. We used basal nutrient solution (BNS)
120 for all hydroponic solutions (Conn *et al.*, 2013). *L. japonicus* seedlings were cultured in
121 5 mL sampling tubes for 2 weeks and transferred to 50 mL sampling tubes with BNS. *M.*
122 *truncatula* seedlings were cultured in 50 mL tubes with BNS from transplant to harvest.
123 For induction of adventitious roots, the roots of 1-week-old seedlings were isolated and
124 cultured in 100 mL of half-strength Gamborg's B5 liquid medium supplemented with
125 3% sucrose and 0.2 mg/L of indole-3-butyric acid. The culture was maintained at 23°C
126 with shaking at 90 rpm and subcultured every 3 to 4 weeks. The *L. japonicus* Lj

suspension cell line (rpc00032) was maintained at 27°C with shaking at 120 rpm in Murashige and Skoog (MS) medium (pH 5.7) supplemented with 1 mg/L of 2,4-D and 3% sucrose, and subcultured every 7 days. For metabolite analysis of cultured cells, ethanol or 100 μM methyl jasmonate (MeJA) was added and the cells were cultured for 4 additional days.

Triterpene extraction and gas chromatography-mass spectrometry (GC-MS) analysis

Triterpenes were extracted from 10 mg of dried plant powder and analyzed by GC-MS following our previous report (Suzuki *et al.*, 2019).

RNA extraction from whole roots

Whole roots (including the hypocotyl) of 2.5-month-old *L. japonicus* cultured on soil or in hydroponics were harvested and immediately frozen in liquid nitrogen. Total RNA extraction, recombinant DNase I treatment, and column purification were performed as described previously (Suzuki *et al.*, 2019).

RNA extraction from SA and SM

Sample fixation, paraffin-embedding, microtome sectioning, collection of SA and SM cell layers by LMD, and RNA extraction were performed following a previous report (Takahashi *et al.*, 2010) with some modifications.

Hypocotyls of *L. japonicus* seedlings cultured hydroponically for 7 weeks were harvested. Segments of hypocotyls were obtained at 5 to 10 mm from the cotyledons. The 5-mm-long hypocotyls were fixed in methanol by incubation for 30 min under a

151 vacuum of 0.09 MPa. This procedure was repeated three times with new methanol.

152 We employed a microwave-assisted method for paraffin-embedding using an
153 Energy Beam Sciences H2850 Microwave Tissue Processor (Agawan, MA). A vial
154 containing fixed tissue and pre-chilled methanol was placed on ice and microwaved
155 three times at 250 W at 37°C for 15 min. The methanol was replaced each time. The
156 temperature in the vial was < 20°C during the first and second microwave treatments;
157 however, 37°C was allowed in the third treatment. The solvent was gradually replaced
158 with 1-butanol, using methanol:1-butanol (3:1), methanol:1-butanol (1:1),
159 methanol:1-butanol (1:3), and 1-butanol, with microwaving at each step at 300 W and
160 37°C for 1.5 min. 1-Butanol was gradually replaced with melted paraffin wax
161 (Surgipath Paraplast X-tra Tissue Infiltration/Embedding Media; Leica Biosystems,
162 Wetzlar, Germany) using 1-butanol:paraffin (3:1), 1-butanol:paraffin (1:1), and
163 1-butanol:paraffin (1:3) with microwaving at each step at 300 W and 58°C for 10 min.
164 Two treatments at 250 W at 58°C for 10 min and five treatments at 250 W at 58°C for
165 30 min were performed with new paraffin solution. The specimens were embedded in
166 paraffin wax and cooled to room temperature. The paraffin blocks were stored at 4°C.

167 RNAsecure reagent (Thermo Fisher Scientific, Waltham, MA) was diluted
168 25-fold with UltraPure Distilled Water (Thermo Fisher Scientific). The diluted
169 RNAsecure solution was placed on an Arcturus PEN Membrane Glass Slide (Thermo
170 Fisher Scientific) heated on a hot plate (57°C). Paraffin sections (12 μm thickness) were
171 floated on the surface of the solution and incubated for several tens of seconds. After
172 removing the solution, the sections were dried at 42°C for 1 h. The slides were
173 immersed in Histoclear II (National Diagnostics, Atlanta, GA) for 10 min twice to
174 remove the paraffin, and air-dried. LMD was performed using a PALM MicroBeam IV

175 Laser-capture Microdissection System (Zeiss, Jena, Germany) with a MicroBeam Laser
176 (cut energy: 55–58, cut focus: 80, LPC energy: 30, LPC focus: 2). AdhesiveCap 500
177 clear (Zeiss) was used to collect cell layers cut by the laser. RNA was purified using an
178 Arcturus PicoPure RNA Isolation Kit (Thermo Fisher Scientific).
179
180 **Quantitative reverse transcription-PCR (qRT-PCR) analysis**
181 First-strand cDNA was synthesized from purified total RNA using PrimeScript RT
182 Master Mix (Perfect Real Time; Takara Bio, Shiga, Japan) following the manufacturer's
183 instructions. The reaction time was 15 and 60 min, for total RNA of whole roots and
184 that of LMD samples, respectively. We performed qRT-PCR analysis following a
185 previous report (Suzuki *et al.*, 2019) using primer nos. 1–20 in Table S1.
186
187 **RNA-seq**
188 Transcriptome profiling of soil-grown and hydroponic roots was performed with three
189 biological replicates (Table S2). Complementary DNA library construction was
190 performed using a TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA)
191 according to the manufacturer's instructions. The resulting cDNA libraries were
192 sequenced using HiSeq 2500 (Illumina) with 100 bp single-end (SE) reads. Adapter
193 sequences and low-quality reads were removed from raw RNA-seq reads by
194 Trimmomatic v. 0.38 (Bolger *et al.*, 2014). Quantification of transcripts was conducted
195 with Salmon v. 0.14.2 (Patro *et al.*, 2017). Reference *L. japonicus* coding sequences
196 (predicted from genome assembly build v. 1.2 of Gifu B-129 accession; Kamal *et al.*,
197 2020) were downloaded from *Lotus* Base (https://lotus.au.dk/). A mapping-based index
198 was built from the reference using an auxiliary k-mer hash over k-mers of length 31 as

recommended. Summarized gene counts were obtained from the transcript counts using the R package tximport v. 1.16.1 (Soneson *et al.*, 2015). Differentially expressed genes (DEGs) were identified using the graphical user interface for TCC (TCC-GUI) (Su *et al.*, 2019). Low count genes (0 mapped reads) were filtered out from TCC computation and default settings were kept for the other parameters.

Transcriptome profiling of the LMD samples was performed with four biological replicates. RNA-seq libraries were constructed and sequenced using a Novaseq 6000 (Illumina) with 150-bp paired-end (PE) reads by Takara Bio. Because two of four replicates of SA samples had smaller total read numbers, they were combined into one FastQ file and analyzed with three biological replicates (Table S2). Data processing was conducted as described above after removing unpaired reads.

Gene cloning

The primers used are listed in Table S1. PCR was performed using Prime STAR MAX or Prime STAR GXL DNA Polymerase (Takara Bio). The plasmid pENTR1A (Thermo Fisher Scientific) was linearized by PCR (primer nos. 21 and 22). The full coding sequences of *LjbHLH32* and *LjbHLH50* were amplified from cDNA of hydroponically cultured roots using primer nos. 24–27 and were integrated into pENTR1A using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, Ipswich, MA) to obtain gateway entry clones (pENTR1A-*LjbHLH32* and pENTR1A-*LjbHLH50*). *GFP* was amplified from pUB-GWS-GFP (GenBank accession no. AB303066) using primer nos. 28 and 29, and cloned into pENTR1A to obtain pENTR1A-*GFP*.

222 **Construction of the gateway destination vector pSD11**

223 The binary vector pUB-GW-GFP is used for overexpression in *L. japonicus* hairy roots
224 (Maekawa *et al.*, 2008). After constructing pUB-GW-GFP, the *Arabidopsis thaliana*
225 heat shock protein (HSP) 18.2 terminator, which is suitable for gene overexpression in
226 plants, was identified (Nagaya *et al.*, 2010; Matsui *et al.*, 2014). Here, we constructed
227 the binary vector pSD11 by replacing the terminators of pUB-GW-GFP with the HSP
228 terminator (Fig. S1).

229 The primers used are listed in Table S1. The 875 bp native *AtHSP18.2*
230 terminator (HSPter) was amplified from *A. thaliana* Col-0 genomic DNA using primer
231 nos. 30 and 31, and cloned into pJET1.2 using a CloneJET PCR Cloning Kit (Thermo
232 Fisher Scientific). The vector backbone was amplified from pUB-GWS-GFP using
233 primer nos. 32 and 33. The HSP terminator was amplified from pJET1.2-HSPter using
234 primer nos. 34 and 35. These two fragments were assembled by NEBuilder HiFi DNA
235 Assembly Master Mix (New England Biolabs) to construct pSD8.

236 The vector backbone, including the gateway cassette and GFP, and fragments,
237 including the *LjUBQ1* promoter, were amplified from pUB-GWS-GFP using primer sets
238 33/36 and 38/39, respectively. These two fragments were assembled with the HSP
239 terminator fragment (primer nos. 34 and 37) to produce pSD9.

240 The gateway destination vector pSD11 was constructed from PCR-amplified
241 fragments of pSD8 (primer nos. 40 and 41) and pSD9 (primer nos. 42 and 43) by the
242 golden gate cloning method using BsaI-HF v. 2 and T4 DNA ligase (both New England
243 Biolabs).

244 **Generation and analysis of TF-overexpressing hairy roots**

245 The expression vectors (pSD11-*GFP*, *LjbHLH32*, and *LjbHLH50*) were generated by

246 LR reaction of pSD11 and entry clones (pENTR1A-*GFP*, *LjbHLH32*, and *LjbHLH50*)
247 using Gateway LR Clonase II Enzyme Mix (Thermo Fisher Scientific). These three
248 vectors were introduced into *Agrobacterium rhizogenes* ATCC15834 by electroporation.
249 *Agrobacterium rhizogenes*-mediated hairy root transformation was performed
250 following our previous study (Suzuki *et al.*, 2019), with some modifications. The
251 incubation temperature was maintained at 25°C in the germination, co-culture, and hairy
252 root induction steps. Isolated hairy roots were cultured at 23°C. For liquid culture of
253 hairy roots, the rotary shaker agitation speed was set to 90 rpm. The
254 germination/co-cultivation medium contained 1/2 B5 basal salts. Hairy root elongation
255 (HRE) medium is composed of 1/2 B5 basal salts, 1/4 Gamborg's B5 vitamins, and 300
256 mg/L of cefotaxime sodium. All media were adjusted to pH 5.5 before autoclaving and
257 freshly prepared. Seeds were germinated as described above. After infection and
258 co-cultivation (Suzuki *et al.*, 2019), seedlings were placed on HRE agar supplemented
259 with 1% sucrose and a filter paper strip was placed on the hypocotyls to prevent the
260 hairy roots from drying. Hairy roots were elongated under a 16 h light/8 h dark cycle for
261 3 weeks, followed by shoot removal. Isolated hypocotyls possessing hairy roots were
262 incubated on new HRE agar containing 3% sucrose without light for 10 days. Single
263 hairy roots were isolated and cultured in 5 mL of liquid HRE medium supplemented
264 with 3% sucrose using six-well cell culture plates. They were subcultured twice every
265 16 days. Hairy root lines emitting GFP fluorescence were selected and each was
266 subcultured into six wells. After 26 days of culture, hairy roots were harvested for
267 GC-MS, qRT-PCR, and RNA-seq analysis (150 bp PE reads, Novaseq 6000 platform,
268 by Takara Bio; Table S2) as described above.
269

270 **Cloning of promoter sequences**

271 Putative promoter sequences, including the 5'-untranslated region (2–2.5 kb upstream
272 from the start codon), of triterpene biosynthetic genes were retrieved using the sequence
273 retrieval tool in *Lotus* Base (https://lotus.au.dk/; Table S3). The primers used to amplify
274 the promoter sequences are listed in Table S1 (primer nos. 44–67). The plasmid
275 pENTR1A (Thermo Fisher Scientific) was linearized by PCR (primer nos. 21 and 23).
276 The promoter sequences of *OSC1*, *OSC3*, *OSC9*, *OSC5*, *CYP716A51*, *LjCYP93E1*,
277 *LjCYP72A61*, *LjCSyGT*, putative 2,3-dihydro-3,5-dihydroxy-6-methyl-4*H*-pyran-4-one
278 (DDMP) transferase (*LjDDMPT*), HMG-CoA reductase 1 (*LjHMGR1*), and squalene
279 epoxidase 1 (*LjSQE1*) were integrated into pENTR1A by NEBuilder HiFi DNA
280 assembly cloning.

281

282 **Vector construction for transient effector-reporter analysis**

283 The *Sfi*I-digested PCR fragment, including the gateway cassette (primer nos. 68 and 69
284 in Table S1) and *Sfi*I-digested pGL4.1HSP [encoding firefly luciferase (fLUC); Nakata
285 *et al.*, 2013], were ligated using T4 DNA ligase (New England Biolabs) to generate the
286 pGL4.1HSP-GW vector. Promoter sequences were integrated into pGL4.1HSP-GW by
287 the LR reaction to generate 11 reporter constructs (pGL4.1HSP-GW-*OSC1pro*,
288 *OSC3pro*, *OSC9pro*, *OSC5pro*, *CYP716A51pro*, *LjCYP93E1pro*, *LjCYP72A61pro*,
289 *LjCSyGTpro*, *LjDDMPTpro*, *LjHMGR1pro*, and *LjSQE1pro*).
290 pENTR-D-*GuCYP93E3pro* (Tamura *et al.*, 2018) was used to generate
291 pGL4.1HSP-GW- *GuCYP93E3pro*.

292 TF genes were transferred from entry clones [pENTR-D-*GubHLH3* (Tamura *et*
293 *al.*, 2018), pENTR1A-*LjbHLH32*, and pENTR1A-*LjbHLH50*] to pDEST35SHSP

294 (Oshima *et al.*, 2013) to produce three effector constructs (pDEST35SHSP-*GubHLH3*,
295 *LjbHLH32*, and *LjbHLH50*). The plasmid pDEST35SHSP-*vesicle-associated membrane*
296 *protein 722* (*Vamp722*; Yoshida *et al.*, 2013) was used as a negative control.
297 We employed phRLHSP [encoding a modified *Renilla* luciferase (RLUC);
298 Yoshida *et al.*, 2013] as an internal reference to normalize the results of our dual
299 luciferase reporter assays.
300
301 **Transient effector-reporter analysis**
302 *A. thaliana* Col-0 was grown under a 16 h light/8 h dark cycle at 22°C for 3–4 weeks.
303 Rosette leaves were used for protoplast isolation after peeling by the Tape-*Arabidopsis*
304 Sandwich method (Wu *et al.*, 2009). Protoplasts were isolated as described (Yoshida *et*
305 *al.*, 2013) with minor modifications. Peeled leaves were incubated for 1 h in enzyme
306 solution [400 mM mannitol, 20 mM KCl, 10 mM $CaCl_2$, 10 g/L of cellulase onozuka
307 R10 (Yakult Pharmaceutical Industry Co., Ltd., Tokyo, Japan), 2.5 g/L of macerozyme
308 R10 (Yakult Pharmaceutical Industry Co., Ltd.), 10 mM 2-mercaptoethanol, and 20 mM
309 2-morpholinoethanesulfonic acid (MES), pH 5.7] in a shaker (22°C, 60 rpm). The
310 protoplasts were filtered through a 70 μm cell strainer and collected by centrifugation at
311 100 × *g* for 10 min. The supernatant was removed by decantation and the remaining
312 pellet was washed thrice with W5 buffer [150 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl,
313 and 2 mM MES, pH 5.7] followed by centrifugation at 100 × *g* for 5 min. The pellet
314 was washed with MMg solution [400 mM mannitol, 15 mM $MgCl_2$, and 4 mM MES,
315 pH 5.7] and resuspended in 5 mL of MMg solution (when using 24 leaves).
316 Reporter, effector, and reference plasmids were co-transfected into protoplasts
317 by the polyethylene glycol (PEG)-assisted method (Yoshida *et al.*, 2013). After

incubating transfected protoplasts at 22°C for 16–19 h, FLUC and RLUC activities were measured using a Pikka Gene Dual Assay Kit (Toyo Ink, Tokyo, Japan).

Phylogenetic analysis

Sequences were collected as described previously (Suzuki *et al.*, 2021). The full-length peptide sequence of CjbHLH1 (BAJ40865.1; Yamada *et al.*, 2011) was retrieved from GenBank (https://www.ncbi.nlm.nih.gov/genbank/). The sequences of Bl (Csa5g156220) and Bt (Csa5g157230; Shang *et al.*, 2014) were obtained from the Cucurbit Genomics Database (http://cucurbitgenomics.org/). A phylogenetic tree was constructed by the neighbor-joining method with 1000 replicates using MEGA6 software (Tamura *et al.*, 2013) following alignment by MUSCLE (Edgar, 2004).

A phylogenetic tree of lateral organ boundaries domain (LBD) proteins was generated using Clustal Omega v. 1.2.3 (Sievers *et al.*, 2011), FastTree v. 2.1.10 (Price *et al.*, 2010), and MEGA X (Kumar *et al.*, 2018) following our previous report (Suzuki *et al.*, 2021).

Results

Hydroponic culture enhances BA accumulation in *L. japonicus* hypocotyl and roots

Previously, we reported triterpene profiling in the model Fabaceae *L. japonicus* grown on soil (Fig. S2; Suzuki *et al.*, 2019). To investigate triterpene production characteristics in this plant, we conducted triterpene sapogenin profiling of whole roots from soil-grown and hydroponically cultured plants, and liquid-cultured adventitious roots (Figs. 1a and S3). Because triterpenes and saponins were extracted and acid-hydrolyzed to remove sugar moieties attached to aglycones, we detected soyasapogenols instead of soyasaponins by GC-MS. BA (peak 6) and soyasapogenol B (SB; a major

342 soyasapogenol in this plant, peak 10) were detected in the extract of soil-grown roots.
343 Liquid culture of adventitious roots induced by indole-3-butyric acid accumulated BA
344 and SB with lower peak intensities. Hydroponic roots of *L. japonicus* accumulated a
345 larger amount of BA than soil-cultured roots and adventitious roots, which caused peak
346 leading in the GC-MS chromatogram. BA was detected in the extracts of roots without
347 acid hydrolysis, indicating that BA is not glycosylated *in planta*. MeJA induces
348 soyasaponin but not BA biosynthesis in Fabaceae plants (Hayashi *et al.*, 2003;
349 Broeckling *et al.*, 2005; Suzuki *et al.*, 2005). We analyzed Lj suspension cultured cells
350 with and without MeJA (Figs. 1b and S3). Neither BA nor SB was detected in the
351 extract of untreated cultured cells. Extracts of the MeJA-treated cells contained a small
352 amount of SB but no BA.

353 We also analyzed the model Fabaceae plant *M. truncatula* (Figs. 1c and S3).
354 Although soil-grown roots of *M. truncatula* (cv. R108) did not accumulate BA, BA was
355 detected in the liquid-cultured adventitious roots and roots of hydroponically cultured
356 plants. To our knowledge, this is the first report of BA accumulation in *Medicago* spp.
357 We also analyzed the adventitious roots of cv. Jemalong A17 and confirmed the
358 presence of BA. However, the amount of BA in *M. truncatula* was much lower than that
359 in *L. japonicus*. *L. japonicus* is more appropriate to study the regulatory mechanisms of
360 BA biosynthesis enhanced in hydroponic culture.

361

362 **SA produces and accumulates BA**

363 Next, we evaluated BA production in hydroponic roots. We analyzed triterpenes
364 extracted from whole roots of *L. japonicus* plants hydroponically cultured for 2–12
365 weeks and observed a time course-dependent increase in BA and enlargement of the

366 hypocotyl and main roots (Fig. S4a–c). Development of a white, spongy tissue was
367 confirmed by anatomical analysis (Fig. 2a), and we assumed that this tissue was
368 homologous to SA in *G. max* (soybean). Soybean develops SA from the SM on stem,
369 hypocotyl, main and lateral roots, and nodules as an adaptive trait to survive flooding by
370 promoting oxygen diffusion in roots (Shimamura *et al.,* 2010; Yamauchi *et al.,* 2013).
371 Hydroponic culture likely resulted in a lack of oxygen in roots, similar to flooding.
372 Hydroponic roots accumulated 10-fold more BA than soil-grown roots, and its quantity
373 reached 6.5 mg/100 mg-dry weight (Fig. 2b). Isolated SA accumulated a higher amount
374 of BA (> 12 mg/100 mg-dry weight) than stele and lateral roots (Fig. 2b). However, SB
375 did not increase in hydroponic roots and SA (Fig. 2c).

376 We analyzed the expression levels of biosynthetic genes for BA (*OSC3* and
377 *CYP716A51*) and SB (*OSC1* and *CYP93E1*) (Fig. 3) by qRT-PCR. When comparing
378 soil-grown and hydroponic roots, BA biosynthetic genes were upregulated in
379 hydroponic roots (Fig. 3). For metabolite analysis (Fig. 2), we harvested SA, including
380 cortical cells. To identify the tissue biosynthesizing BA, we isolated SA and SM cell
381 layers by LMD (Fig. S5) and performed RNA extraction and qRT-PCR. Because BA
382 biosynthetic gene expression was higher in SA (Fig. 3), more BA is likely synthesized
383 in SA. These results suggest that transcriptional regulation is responsible for the
384 increased BA level in SA of *L. japonicus* hydroponic roots.

385

386 **Candidate regulators of BA biosynthesis as identified by RNA-seq**

387 We performed an RNA-seq-based comparative transcriptome analysis for mining
388 potential transcriptional regulators of BA biosynthesis. RNA evaluated by qRT-PCR
389 (Fig. 3) was used for the analysis, and sequencing of the whole root samples and the

390 LMD samples was conducted in three and four replicates, respectively. Data processing
391 was performed in triplicate for all samples as mentioned in Materials and Methods. We
392 performed differential expression analyses independently in data set 1 [soil-grown roots
393 (So) and hydroponic roots (Hy)] (Figs. 4a and S6a) and data set 2 (SM and SA) (Figs.
394 4b and S6b). Based on the MA plot (Fig. S6), we identified 3152 and 1994 DEGs whose
395 expression was higher in So and Hy, respectively (Table S4). SM and SA had 2169 and
396 1647 DEGs with elevated expression levels (Table S4). There were 256 DEGs
397 overlapping between Hy and SA (Fig. 4c), including 23 putative TF genes (Fig. 4c and
398 Table S5), as well as BA biosynthetic genes (*OSC3* and *CYP716A51*) and genes for
399 rate-limiting enzymes for triterpene biosynthesis [3-hydroxy-3-methyl-glutaryl-CoA
400 reductase1 (HMGR1) and squalene epoxidase1 (SQE1)].

401 We assessed the correlation between the expression patterns of BA biosynthetic
402 genes and those of candidate TF genes using the gene expression atlas in *Lotus* Base
403 (https://lotus.au.dk/expat/; 81 conditions based on annotation and prediction in
404 Miyakojima MG-20 genome v. 3.0). Seven TF genes were co-expressed with *OSC3*,
405 *CYP716A51*, *HMGR1*, and *SQE1* (Pearson's correlation coefficient [PCC] threshold:
406 0.4; Table 1). Among the seven TFs, we selected LjbHLH32 (LotjaGi4g1v0043000) and
407 LjbHLH50 (LotjaGi1g1v0340900) for further analysis because bHLH TFs are triterpene
408 biosynthesis regulators in Fabaceae (Mertens *et al.*, 2016; Tamura *et al.*, 2018), *Panax*
409 spp. (Zhang *et al.*, 2017), cucumber (Shang *et al.*, 2014), and *B. platyphylla* (Yin *et al.*,
410 2017).

411 The bHLH TFs were classified into approximately 25 clades based on sequence
412 similarity of the bHLH domain and other conserved domains (Heim *et al.*, 2003). We
413 constructed a phylogenetic tree of full-length bHLH proteins of clades Ia, Ib, and IVa,

414 including representative bHLH proteins involved in specialized plant metabolism (Fig.
415 4d). Fabaceae clade IVa bHLH TFs were further subdivided into three subclades, and
416 subclade IVa1 included all Fabaceae triterpene saponin biosynthesis activators identified
417 in clade IVa (Suzuki *et al.*, 2021). Interestingly, LjbHLH32 was a member of subclade
418 IVa2 but not IVa1 (Fig. 4d). The orthologs in subclade IVa2 were highly conserved in
419 Fabaceae plants and were commonly expressed in roots and nodules of model Fabaceae
420 plants (Suzuki *et al.*, 2021). However, no subclade IVa2 bHLHs have been functionally
421 characterized to date. The closest homolog of another candidate LjbHLH50 in *M.*
422 *truncatula* is MtbHLH1 (Medtr3g099620), which is suggested to regulate nitrogen
423 transfer between nodule and root cells by controlling nodule vascular patterning
424 (Godiard *et al.*, 2011). LjbHLH50 and MtbHLH1 showed 60% amino acid identity
425 (99% coverage) and belonged to clade Ia (Fig. 4d), similar to BpbHLH9, an OA and BA
426 biosynthesis regulator in *B. platyphylla* (Yin *et al.*, 2017). These phylogenetic
427 relationships between our candidates and known triterpene biosynthesis regulators
428 suggest that LjbHLH32 and LjbHLH50 are involved in BA biosynthesis regulation in *L.*
429 *japonicus*.
430
431 **Overexpression of LjbHLH32 or LjbHLH50 enhances BA production in transgenic**
432 **hairy roots**
433 We generated GFP-expressing (control), LjbHLH32-overexpressing (LjbHLH32-OX),
434 and LjbHLH50-OX hairy root lines for functional identification of LjbHLH32 and
435 LjbHLH50 *in planta*. We chose three independent lines for each construct and analyzed
436 their triterpene content by GC-MS (Fig. 5a). Compared to the control lines,
437 LjbHLH32-OX and LjbHLH50-OX hairy roots accumulated 5.73- and 4.31-fold more

438 BA (Fig. 5a, right), but the amount of SB was weakly elevated (Fig. 5a, left). Next, we
439 assessed the expression of triterpene biosynthetic genes by qRT-PCR (Fig. 5b).
440 LjbHLHs-OX lines overexpressed the corresponding bHLH genes. LjbHLH32
441 overexpression increased the transcript levels of the BA biosynthetic genes *LjHMGR1*,
442 *LjSQE1*, *OSC3*, and *CYP716A51*, whereas branch pathway genes [cycloartenol synthase
443 gene (*OSC5*), *OSC1*, and *LjCYP93E1*] were only slightly affected. LjbHLH50-OX lines
444 showed 1.56–2.26-fold higher expression of *LjHMGR1*, *LjSQE1*, *OSC5*, *OSC1*, and
445 *LjCYP93E1*. However, the expression of *OSC3* and *CYP716A51* was enhanced to a
446 greater degree.

447 We performed an RNA-seq analysis of nine hairy root lines (Fig. 5) to test the
448 effect of LjbHLH32 and LjbHLH50 overexpression on a genome-wide scale. The
449 control lines and bHLH-OX lines were distinguished by principal component analysis
450 (PCA) of transcriptome data (Fig. 6a and b). DEGs were identified based on an MA plot
451 (Fig. S6c and d). Compared to the control lines, 527 and 1045 genes were significantly
452 upregulated in LjbHLH32-OX and LjbHLH50-OX, respectively (Table S4). Among the
453 DEGs upregulated in bHLH-OX hairy roots and 256 DEGs commonly upregulated in
454 Hy and SA samples (Fig. 4c), 6 overlapped (Fig. 6c). These six genes included the BA
455 biosynthetic genes *OSC3* and *CYP716A51* but no genes involved in other triterpene
456 biosynthesis pathways (Table S6). Some of the enzyme genes putatively involved in
457 precursor supply for BA biosynthesis were significantly upregulated by overexpression
458 of bHLHs (Table S7), which may contribute to increased BA production in hairy roots
459 (Fig. 5a).

460 The genes upregulated in the LjbHLH32- and LjbHLH50-OX lines (Fig. 6c)
461 are likely involved in the metabolism of primary metabolites and plant hormones,

transport of metabolites, signaling, and defense responses (Table S6). *LjHMGR1* and *LjSQE1* were significantly upregulated by LjbHLH32-overexpression (Table 2). Interestingly, LjbHLH32-OX hairy roots showed significantly elevated expression of genes (*AMY2*, *CYP88D4*, and *CYP88D5*) in the *AMY2* gene cluster (Tables 2 and S7; Krokida *et al.,* 2013). The products of the *AMY2* gene cluster in *L. japonicus* are unknown. AMY2 produced β-amyrin and lupeol in a yeast expression system (Iturbe-Ormaetxe *et al.,* 2003), and β-amyrin and dihydro-lupeol by agro-infiltration into tobacco leaves (Krokida *et al.,* 2013). CYP88D4 and CYP88D5, homologs of CYP88D6 (involved in glycyrrhizin biosynthesis in *G. uralensis*), did not oxidize the tested triterpenes (Seki *et al.,* 2008; Krokida *et al.,* 2013). CYP71D353 showed dihydro-lupeol oxidation activity when transiently expressed in tobacco leaves (Krokida *et al.,* 2013); however, the expression level of *CYP71D353* was not affected by overexpression of LjbHLH32 (Table 2).

LjbHLH50-overexpression resulted in significant upregulation of several of the 23 TF genes selected by RNA-seq (Table 1), but those TF genes were not upregulated by LjbHLH32 overexpression (Table S6).

LjbHLH32 shows similar trends to GubHLH3 in a transient effector-reporter analysis

Analyses of hairy roots indicated that overexpression of LjbHLH32 or LjbHLH50 upregulated BA biosynthesis directly or indirectly. To obtain insight into how they regulate BA biosynthesis, we examined the transactivation abilities of LjbHLH32 and LjbHLH50 against the promoter sequences of triterpene biosynthetic genes (Table S3) by transient effector-reporter analysis. Effector (*35S*::*bHLH*), reporter

486 (*promoter*::*FLUC*), and reference (*35S*::*RLUC*) constructs were co-transfected into
487 *Arabidopsis* mesophyll cell protoplasts and transactivation activities were measured as
488 relative luciferase activities (FLUC/RLUC; Fig. 7). We used the combination of
489 *GubHLH3* and *GuCYP93E3pro* as the positive control and confirmed a similar level of
490 transactivation (39-fold) as a previous transient effector-reporter analysis using tobacco
491 BY-2 cell protoplasts (Tamura *et al.*, 2018). Similarly, GubHLH3 transactivated
492 promoters of soyasaponin biosynthetic genes in *L. japonicus* except *OSC1pro*
493 (*LjCYP93E3pro*, *LjCYP72A61pro*, *LjCSyGTpro*, and *LjDDMPTpro*) and the genes
494 involved in precursor supply (*LjHMGR1pro* and *LjSQE1pro*). GubHLH3 weakly
495 transactivated the promoter of *OSC5* (*cycloartenol synthase*), consistent with a previous
496 study (Tamura *et al.*, 2018). LjbHLH32 and LjbHLH50 did not transactivate *OSC3pro*
497 and *CYP716A51pro*. Unexpectedly, LjbHLH32 significantly transactivated the
498 promoters of genes in the soyasaponin, precursor, and sterol pathways. Although the
499 fold-changes induced by LjbHLH32 were generally lower than those by GubHLH3, the
500 transactivation pattern of LjbHLH32 was highly correlated to that of GubHLH3
501 (PCC = 0.99). LjbHLH50 did not positively affect any promoters and the results were
502 not correlated to those of GubHLH3 (PCC = 0.36).
503

504 **Discussion**

505 **SA formation and BA biosynthesis**

506 Plant specialized metabolites are often produced and accumulated in specific cell types
507 and organs, such as artemisinin in *Artemisia annua* (Duke & Paul, 1993; Duke *et al.*,
508 1994), nicotine in tobacco (Katoh *et al.*, 2005), glycyrrhizin in *Glycyrrhiza* spp.
509 (Kojoma *et al.*, 2010; Li *et al.*, 2014), and monoterpene indole alkaloid (MIA) in

Catharanthus roseus (Yamamoto *et al.*, 2019). In this study, we identified SA as a sink tissue of BA in *L. japonicus*, which is formed on the surface of hypocotyls and roots when cultured hydroponically (Figs. 2 and S4). BA negatively regulates nodulation by controlling *ENOD40* (early nodulin gene) expression in *L. japonicus* (Delis *et al.*, 2011; Suzuki *et al.*, 2019). More BA accumulated in SA than in roots of soil-grown plants (Figs. 1a and 2), suggesting an unknown and SA-specific role of BA. The role of aerenchyma, including SA, is to supply air to roots under flooded conditions (Colmer & Pedersen, 2008; Shimamura *et al.*, 2010; Takahashi *et al.*, 2014). Our recent study revealed that *G. max* also accumulated BA in SA under flooded conditions and suggested that the accumulation of hydrophobic BA was essential to exclude water from gas spaces and promote air diffusion (Takahashi et al. 2022).

Sucrose supply from leaves is required for SA formation (Striker & Colmer, 2017; Takahashi *et al.*, 2018). However, it remains elusive whether sucrose functions only as a nutrient during cell differentiation from SM to SA, or also as a signaling molecule (Tognetti *et al.*, 2013). Hydroponically cultured *L. japonicus* will be useful to investigate such molecular mechanisms of SA formation.

Triterpene biosynthesis is downregulated in underground tissues of *Glycyrrhiza* spp. lacking leaves during winter (Hayashi *et al.*, 1998; Ramilowski *et al.*, 2013), and the total amount of triterpene decreased in aseptic liquid cultures of *G. uralensis* stolons lacking aerial parts (Kojoma *et al.*, 2010; Tamura *et al.*, 2017). Liquid-cultured *L. japonicus* adventitious roots also showed decreased triterpene contents compared to the roots of soil-grown plants (Fig. 1a). Photosynthesis and sucrose supply from aerial parts may affect not only SA formation but also triterpene biosynthesis in underground tissues of Fabaceae plants.

534

535 **Subclade IVa1 and IVa2 bHLHs are likely functionally diverged in Fabaceae**
536 **triterpene biosynthesis regulation**

537 Clade IVa bHLHs are conserved factors regulating jasmonate-inducible terpene
538 metabolism in plants (Goossens *et al.*, 2017; da Silva Magedans *et al.*, 2020), including
539 that of MIA in *C. roseus* (Van Moerkercke *et al.*, 2015, 2016), and triterpene saponins
540 in *Chenopodium quinoa* (Jarvis *et al.*, 2017), *Panax notoginseng* (Zhang *et al.*, 2017),
541 and Fabaceae plants (Mertens *et al.*, 2016; Tamura *et al.*, 2018; Ribeiro *et al.*, 2020).
542 We showed that clade IVa bHLHs in Fabaceae are subdivided into three subclades; all
543 saponin biosynthesis activators identified to date in Fabaceae fell into subclade IVa1
544 (Suzuki *et al.*, 2021). BA biosynthesis is jasmonate-independent in Fabaceae plants
545 (Hayashi *et al.*, 2003; Broeckling *et al.*, 2005; Suzuki *et al.*, 2005), and LjbHLH32, a
546 member of subclade IVa2, activated BA biosynthesis in *L. japonicus* transgenic hairy
547 roots (Fig. 5). These results suggest that subclade IVa2 gained new target specificities
548 during the evolution from subclade IVa1 in Fabaceae plants.

549 Interestingly, similar to the soyasaponin biosynthesis activator GubHLH3,
550 LjbHLH32 transactivated the promoters of soyasaponin biosynthetic genes but not of
551 BA biosynthetic genes in transient effector-reporter assays using *A. thaliana* mesophyll
552 cell protoplasts (Fig. 7). The changes in *cis*-elements in promoters and DNA-binding
553 domains in TFs are likely important to rewire gene regulatory networks (Shoji, 2019).
554 In the case of MtTSARs, GubHLH3, and LjbHLH32, however, peptide sequences of the
555 basic domains are well conserved (Fig. S7). Differences in protein-protein interactions
556 among TFs may contribute to changes in target recognition by TFs. The bHLH-MYB
557 complex is implicated in the transcriptional regulation of anthocyanin biosynthesis. In

558 kiwifruit, *Actinidia chinensis*, the complex of AcbHLH42 and AcMYB123 regulated
559 tissue-specific anthocyanin biosynthesis (Wang *et al.,* 2019). Although single
560 overexpression of AcbHLH42 or AcMYB123 in *Arabidopsis* was not enough to
561 increase the anthocyanin content, their co-expression induced anthocyanin production in
562 seedlings (Wang *et al.,* 2019). LjbHLH32 may have missed its interacting partners in
563 *Arabidopsis* mesophyll cells (Fig. 7), unlike *L. japonicus* hairy roots (Fig. 5). To prove
564 this hypothesis, it is necessary to identify direct targets and interacting partners of
565 LjbHLH32.
566
567 **Conserved roles of clade Ia bHLH TFs in BA biosynthesis regulation**
568 Clade Ia bHLHs are subdivided into the SMF subclade (described in detail later)
569 regulating stomatal development and the others subclade (Ran *et al.,* 2013). BpbHLH9
570 in the others subclade is an OA and BA biosynthesis regulator in *B. platyphylla* (Yin *et*
571 *al.,* 2017). LjbHLH50 is also of the others subclade and enhanced BA biosynthesis in
572 *Lotus* transgenic hairy roots (Fig. 5), suggesting that bHLHs in this subclade are
573 conserved regulators of BA biosynthesis in fabids.
574 LjbHLH50 did not transactivate the promoters tested (Fig. 7), suggesting that it
575 lacked its interacting partners in *Arabidopsis* mesophyll cells or indirectly regulated BA
576 biosynthesis via downstream TF cascades. Although overexpression of BpbHLH9
577 enhanced expression of some triterpene biosynthesis genes and accumulation of
578 triterpene compounds, the genes directly regulated by BpbHLH9 are unknown (Yin *et*
579 *al.,* 2017). Our first transcriptome analysis of whole roots from soil-grown and
580 hydroponically cultured plants, and SA and SM cell layers isolated by LMD revealed 23
581 candidate TFs, including LjbHLH32 and LjbHLH50 (Fig. 4 and Table 1). We also

582 performed transcriptome profiling of transgenic hairy root lines (Fig. 6). We found
583 upregulation of 6 of 23 TF candidate genes by overexpression of LjbHLH50, which
584 were annotated as LBD, ERF, NAC, and SCREAM (SCRM)-like proteins (Table S6).
585 LjbHLH50 (and BpbHLH9) may indirectly regulate BA biosynthesis by upregulating
586 these TF genes.

587 In higher plants, SPEECHLESS (SPCH), MUTE, and FAMA in the SMF
588 subclade regulate stomatal development (Ohashi-Ito & Bergmann, 2006; MacAlister *et*
589 *al.*, 2007; Pillitteri *et al.*, 2007). The clade IIIb bHLHs SCRM (ICE1) and SCRM2 are
590 essential for stomatal development because they formed heterodimers with SPCH,
591 MUTE, and FAMA (Kanaoka *et al.*, 2008). AtSCRM protein (At3g26744.1; 494 aa) has
592 bHLH (298–350 aa) and putative zipper (ZIP; 351–380 aa) domains (Chinnusamy *et al.*,
593 2003). Alignments between AtSCRM and AtSCRM-like (At2g40435.1) proteins
594 showed that SCRM-like proteins lacked a bHLH domain but possessed a partial ZIP
595 domain (Fig. S8). An analysis of the interaction between LjbHLH50 (BpbHLH9) and
596 SCRM-like proteins is interesting.

597

598 **Sleeping BA biosynthesis ability in *M. truncatula***

599 The *M. truncatula* genome has putative lupeol synthase genes but lupeol derivatives
600 have not been detected in *Medicago* spp. (Naoumkina *et al.*, 2010). We found that
601 whole roots from hydroponically cultured plants and liquid-cultured adventitious roots
602 accumulated trace amounts of BA (Fig. 1c). Because homologs of LjbHLH32 and
603 LjbHLH50 were present in *M. truncatula* (Fig. 4d), expression of these TFs might limit
604 BA biosynthesis in this plant.

605 Silencing of *OSC3* expression and reduced accumulation of lupeol derivatives

results in upregulated *ENOD40* expression and enhanced nodulation in *L. japonicus* (Delis *et al.*, 2011). Similarly, in *M. truncatula* roots, overexpression of chimeric MtbHLH1 fused to the EAR repression domain (transgenic plants should show dominant loss-of-function phenotypes; Hiratsu *et al.*, 2003) resulted in abnormal nodule vascular bundle development and slightly increased numbers of nodules per plant although the growth of untransformed aerial parts was significantly reduced in the composite plants (Godiard *et al.*, 2011). Although it is unknown whether MtbHLH1 positively regulates BA biosynthesis similar to its ortholog LjbHLH50 and whether *M. truncatula* nodules accumulate lupeol derivatives, Delis *et al.* (2011) and Godiard *et al.* (2011) likely reported similar phenomena. Re-evaluation of these mutants will provide rigid understanding for the role of lupeol derivatives and clade Ia bHLHs in nodule formation.

Godiard *et al*. (2011) suggested that an LBD-type TF gene (Medtr7g007010) is a potential target of MtbHLH1 by transcriptome analysis of transgenic roots overexpressing MtbHLH1-EAR. An LBD gene was co-upregulated in hydroponic roots, SA, and LjbHLH50-OX hairy roots (Lj5g3v2099760/LotjaGi5g1v0323400; Table S6) and closely related to Medtr07g007010 in the phylogenetic tree (Fig. S9), suggesting the involvement of LBD proteins in this yet-to-be-defined TF cascade.

Concluding remarks

We identified two bHLH-type TFs enhancing BA biosynthesis by transcriptome analysis of hydroponically cultured *L. japonicus* and a functional analysis using transgenic hairy roots. However, neither LjbHLH32 nor LjbHLH50 transactivated the promoters of BA biosynthetic genes in transient effector-reporter assays using a heterologous host

630 (*Arabidopsis*), suggesting that these bHLHs interact with unknown TFs or indirectly
631 regulate BA biosynthesis via downstream TFs. The 23 putative TFs found in the first
632 screening are promising targets for further analysis (Table 1). This study is a first step
633 toward understanding the complicated transcriptional regulatory mechanisms of BA
634 biosynthesis in *L. japonicus*.
635
636 **Data Availability**
637 The nucleotide sequence reported in this paper has been submitted to the DNA Data
638 Bank of Japan (DDBJ) under the accession numbers of LC609881 (*LjbHLH32*) and
639 LC609882 (*LjbHLH50*). The raw RNA-seq reads obtained were deposited in the DDBJ
640 Sequence Read Archive (DRA) under the accession number DRA011635.
641
642 **Acknowledgements**
643 We thank Miyazaki University and National BioResource Project (NBRP) for providing
644 pUB-GWS-GFP plasmid and *L. japonicus* Gifu B-129 seeds, RIKEN BRC for
645 providing Lj suspension cell line (rpc00032) through NBRP, the Institut National de la
646 Recherche Agronomique for providing *M. truncatula R108* and Jemalong A17 seeds,
647 Prof. Shuhei Yasumoto (Osaka University) for providing genomic DNA of *A. thaliana*
648 Col-0, and Dr. Atsushi Toyoda (National Institute of Genetics) for construction and
649 sequencing of RNA-seq library of soil-grown and hydroponically-cultured whole roots.
650 We also thank Dr. Nobutaka Mitsuda, Dr. Shingo Sakamoto and Ms. Tomoko Niki
651 (AIST) for technical support on transient effector-reporter analysis. This work was
652 supported by the Ministry of Education, Culture, Sports, Science and Technology
653 (MEXT), Japan [JSPS KAKENHI grant Nos. 16H06279 (PAGS), 17K07754 and

JP20H02913 to HSe, JP19H02921 to TM and JP19J10245 to HSu, and the Frontier Research Base for Global Young Researchers, Osaka University, to EOF].

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contribution

EOF, TM and HSe conceived and supervised the study; All authors designed the experiments; HSu and HT performed the experiments; HSu wrote the manuscript; HT, EOF, MN, TM and HSe made manuscript revisions. All authors read and approved the final manuscript.

References

Bolger AM, Lohse M, Usadel B. **2014**. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Broeckling CD, Huhman DV, Farag MA, Smith JT, May GD, Mendes P, Dixon RA, Sumner LW. **2005**. Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. *Journal of Experimental Botany* **56**: 323–336.

Chinnusamy V, Ohta M, Kanrar S, Lee B, Hong X, Agarwal M, Zhu J-K. **2003**. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. *Genes & Development* **17**: 1043–1054.

Chung SY, Seki H, Fujisawa Y, Shimoda Y, Hiraga S, Nomura Y, Saito K, Ishimoto M, Muranaka T. **2020**. A cellulose synthase-derived enzyme catalyses 3- *O* -glucuronosylation in saponin biosynthesis. *Nature Communications* **11**: 5664.

Colmer TD, Pedersen O. **2008**. Oxygen dynamics in submerged rice (*Oryza sativa*).

New Phytologist **178**: 326–334.

Conn SJ, Hocking B, Dayod M, Xu B, Athman A, Henderson S, Aukett L, Conn V, Shearer MK, Fuentes S, *et al.* 2013. Protocol: optimising hydroponic growth systems for nutritional and physiological analysis of *Arabidopsis thaliana* and other plants. *Plant Methods* **9**: 4.

da Silva Magedans YV, Phillips MA, Fett-Neto AG. **2020**. Production of plant bioactive triterpenoid saponins: from metabolites to genes and back. *Phytochemistry Reviews* **20**: 461–482.

Delis C, Krokida A, Georgiou S, Peña-Rodríguez LM, Kavroulakis N, Ioannou E, Roussis V, Osbourn AE, Papadopoulou KK. **2011**. Role of lupeol synthase in *Lotus japonicus* nodule formation. *New Phytologist* **189**: 335–346.

Duke SO, Paul RN. **1993**. Development and fine structure of the glandular trichomes of *Artemisia annua* L. *International Journal of Plant Sciences* **154**: 107–118.

Duke MV, Paul RN, Elsohly HN, Sturtz G, Duke SO. **1994**. Localization of artemisinin and artemisitene in foliar tissues of glanded and glandless biotypes of *Artemisia annua* L. *International Journal of Plant Sciences* **155**: 365–372.

Edgar RC. **2004**. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Fenwick G. **1992**. Saponins. *Toxic substances in crop plants*: 285–327.

Godiard L, Lepage A, Moreau S, Laporte D, Verdenaud M, Timmers T, Gamas P. **2011**. MtbHLH1, a bHLH transcription factor involved in *Medicago truncatula* nodule vascular patterning and nodule to plant metabolic exchanges. *New Phytologist* **191**: 391–404.

Goossens J, Mertens J, Goossens A. **2017**. Role and functioning of bHLH transcription factors in jasmonate signalling. *Journal of Experimental Botany* **68**: 1333–1347.

Hayashi H, Hiraoka N, Ikeshiro Y, Yamamoto H, Yoshikawa T. **1998**. Seasonal variation of glycyrrhizin and isoliquiritigenin glycosides in the root of *Glycyrrhiza glabra* L. *Biological & Pharmaceutical Bulletin* **21**: 987–989.

Hayashi H, Huang P, Inoue K. **2003**. Up-regulation of soyasaponin biosynthesis by methyl jasmonate in cultured cells of *Glycyrrhiza glabra*. *Plant and Cell Physiology* **44**: 404–411.

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. 2003. The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. *Molecular Biology and Evolution* **20**: 735–747.

Hiratsu K, Matsui K, Koyama T, Ohme-Takagi M. **2003**. Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *The Plant Journal* **34**: 733–739.

Horacio P, Martinez-Noel G, Tognetti JA. **2013**. Sucrose signaling in plants: A world yet to be explored. *Plant Signaling & Behavior* **8**: e23316.

Huhman DV, Berhow MA, Sumner LW. **2005**. Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*. *Journal of Agricultural and Food Chemistry* **53**: 1914–1920.

Iturbe-Ormaetxe I, Haralampidis K, Papadopoulou K, Osbourn AE. **2003**. Molecular cloning and characterization of triterpene synthases from *Medicago truncatula* and *Lotus japonicus*. *Plant Molecular Biology* **51**: 731–743.

Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, Ohyanagi H, Mineta K, Michell CT, Saber N, *et al.* 2017. The genome of *Chenopodium quinoa*. *Nature* **542**: 307–312.

Jozwiak A, Sonawane PD, Panda S, Garagounis C, Papadopoulou KK, Abebie B, Massalha H, Almekias-Siegl E, Scherf T, Aharoni A. **2020**. Plant terpenoid metabolism co-opts a component of the cell wall biosynthesis machinery. *Nature Chemical Biology* **16**: 740–748.

Kamal N, Mun T, Reid D, Lin JS, Akyol TY, Sandal N, Asp T, Hirakawa H, Stougaard J, Mayer KFX, Sato S, Andersen SU. 2020. Insights into the evolution of symbiosis gene copy number and distribution from a chromosome-scale *Lotus japonicus* Gifu genome sequence. *DNA Research* **27**:dsaa015.

Kanaoka MM, Pillitteri LJ, Fujii H, Yoshida Y, Bogenschutz NL, Takabayashi J,

Zhu J-K, Torii KU. **2008**. *SCREAM/ICE1* and *SCREAM2* specify three cell-state transitional steps leading to *Arabidopsis* stomatal differentiation. *The Plant Cell* **20**: 1775–1785.

Katoh A, Ohki H, Inai K, Hashimoto T. **2005**. Molecular regulation of nicotine biosynthesis. *Plant Biotechnology* **22**: 389–392.

Kojoma M, Ohyama K, Seki H, Hiraoka Y, Asazu SN, Sawa S, Sekizaki H, Yoshida S, Muranaka T. **2010**. *In vitro* proliferation and triterpenoid characteristics of licorice (*Glycyrrhiza uralensis* Fischer, Leguminosae) stolons. *Plant Biotechnology* **27**: 59–66.

Krokida A, Delis C, Geisler K, Garagounis C, Tsikou D, Peña-Rodríguez LM, Katsarou D, Field B, Osbourn AE, Papadopoulou KK. **2013**. A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis. *New Phytologist* **200**: 675–690.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. **2018**. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**: 1547–1549.

Li B, Bhandari DR, Janfelt C, Römpp A, Spengler B. **2014**. Natural products in *Glycyrrhiza glabra* (licorice) rhizome imaged at the cellular level by atmospheric pressure matrix-assisted laser desorption/ionization tandem mass spectrometry imaging. *The Plant Journal* **80**: 161–171.

MacAlister CA, Ohashi-Ito K, Bergmann DC. **2007**. Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. *Nature* **445**: 537–540.

Maekawa T, Kusakabe M, Shimoda Y, Sato S, Tabata S, Murooka Y, Hayashi M. **2008**. Polyubiquitin promoter-based binary vectors for overexpression and gene silencing in *Lotus japonicus*. *Molecular Plant-Microbe Interactions* **21**: 375–382.

Matsui T, Sawada K, Takita E, Kato K. **2014**. The longer version of *Arabidopsis thaliana heat shock protein 18.2* gene terminator contributes to higher expression of stably integrated transgenes in cultured tobacco cells. *Plant Biotechnology* **31**: 191–194.

Mertens J, Pollier J, Bossche RV, Lopez-Vidriero I, Franco-Zorrilla JM, Goossens A. **2016**. The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene

saponin biosynthesis in *Medicago truncatula*. *Plant Physiology* **170**: 194–210.

Mochizuki T, Takahashi U, Shimamura S, Fukuyama M. **2000**. Secondary aerenchyma formation in hypocotyl in summer leguminous crops. *Japanese Journal of Crop Science* **69**: 69–73.

Nagaya S, Kawamura K, Shinmyo A, Kato K. **2010**. The HSP terminator of *Arabidopsis thaliana* increases gene expression in plant cells. *Plant and Cell Physiology* **51**: 328–332.

Naoumkina MA, Modolo LV, Huhman DV, Urbanczyk-Wochniak E, Tang Y, Sumner LW, Dixon RA. **2010**. Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. *The Plant Cell* **22**: 850–866.

Ohashi-Ito K, Bergmann DC. **2006**. *Arabidopsis* FAMA controls the final proliferation/differentiation switch during stomatal development. *The Plant Cell* **18**: 2493–2505.

Oshima Y, Shikata M, Koyama T, Ohtsubo N, Mitsuda N, Ohme-Takagi M. **2013**. MIXTA-Like transcription factors and WAX INDUCER1/SHINE1 coordinately regulate cuticle development in *Arabidopsis* and *Torenia fournieri*. *The Plant Cell* **25**: 1609–1624.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. **2017**. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**: 417–419.

Pillitteri LJ, Sloan DB, Bogenschutz NL, Torii KU. **2007**. Termination of asymmetric cell division and differentiation of stomata. *Nature* **445**: 501–505.

Price MN, Dehal PS, Arkin AP. **2010**. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**: e9490.

Ramilowski JA, Sawai S, Seki H, Mochida K, Yoshida T, Sakurai T, Muranaka T, Saito K, Daub CO. **2013**. *Glycyrrhiza uralensis* transcriptome landscape and study of phytochemicals. *Plant and Cell Physiology* **54**: 697–710.

Ran J-H, Shen T-T, Liu W-J, Wang X-Q. **2013**. Evolution of the bHLH genes involved in stomatal development: implications for the expansion of developmental

complexity of stomata in land plants. *PLoS ONE* **8**: e78997.

Ribeiro B, Lacchini E, Bicalho KU, Mertens J, Arendt P, Bossche RV, Calegario G, Gryffroy L, Ceulemans E, Buitink J, *et al.* 2020. A seed-specific regulator of triterpene saponin biosynthesis in *Medicago truncatula*. *The Plant Cell* **32**: 2020–2042.

Seki H, Ohyama K, Sawai S, Mizutani M, Ohnishi T, Sudo H, Akashi T, Aoki T, Saito K, Muranaka T. **2008**. Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proceedings of the National Academy of Sciences* **105**: 14204–14209.

Seki H, Tamura K, Muranaka T. **2015**. P450s and UGTs: Key players in the structural diversity of triterpenoid saponins. *Plant and Cell Physiology* **56**: 1463–1471.

Shang Y, Ma Y, Zhou Y, Zhang H, Duan L, Chen H, Zeng J, Zhou Q, Wang S, Gu W, *et al.* 2014. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* **346**: 1084–1088.

Shimamura S, Yamamoto R, Nakamura T, Shimada S, Komatsu S. **2010**. Stem hypertrophic lenticels and secondary aerenchyma enable oxygen transport to roots of soybean in flooded soil. *Annals of Botany* **106**: 277–284.

Shoji T. **2019**. The recruitment model of metabolic evolution: jasmonate-responsive transcription factors and a conceptual model for the evolution of metabolic pathways. *Frontiers in Plant Science* **10**: 560.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, *et al.* 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**: 539.

Soneson C, Love MI, Robinson MD. **2015**. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**: 1521.

Striker GG, Colmer TD. **2017**. Flooding tolerance of forage legumes. *Journal of Experimental Botany* **68**: 1851–1872.

Su W, Sun J, Shimizu K, Kadota K. **2019**. TCC-GUI: a Shiny-based application for differential expression analysis of RNA-Seq count data. *BMC Research Notes* **12**: 133.

Suzuki H, Fukushima EO, Shimizu Y, Seki H, Fujisawa Y, Ishimoto M, Osakabe K,

Osakabe Y, Muranaka T. **2019**. *Lotus japonicus* triterpenoid profile and characterization of the *CYP716A51* and *LjCYP93E1* genes involved in their biosynthesis in planta. *Plant and Cell Physiology* **60**: 2496–2509.

Suzuki H, Reddy MSS, Naoumkina M, Aziz N, May GD, Huhman DV, Sumner LW, Blount JW, Mendes P, Dixon RA. **2005**. Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures of the model legume *Medicago truncatula*. *Planta* **220**: 696–707.

Suzuki H, Seki H, Muranaka T. **2021**. Insights into the diversification of subclade IVa bHLH transcription factors in Fabaceae. *BMC Plant Biology* **21**: 109.

Takahashi H, Abo C, Suzuki H, Romsuk J, Oi T, Yanagawa A, Gorai T, Tomisaki Y, Jitsui M, Shimamura S *et al*. 2022. Triterpenoids in aerenchymatous phellem contribute to internal root aeration and waterlogging adaptability in soybean. *Research Square*. https://doi.org/10.21203/rs.3.rs-2230730/v1

Takahashi H, Kamakura H, Sato Y, Shiono K, Abiko T, Tsutsumi N, Nagamura Y, Nishizawa NK, Nakazono M. **2010**. A method for obtaining high quality RNA from paraffin sections of plant tissues by laser microdissection. *Journal of Plant Research* **123**: 807–813.

Takahashi H, Xiaohua Q, Shimamura S, Yanagawa A, Hiraga S, Nakazono M. **2018**. Sucrose supply from leaves is required for aerenchymatous phellem formation in hypocotyl of soybean under waterlogged conditions. *Annals of Botany* **121**: 723–732.

Takahashi H, Yamauchi T, Colmer TD, Nakazono M. **2014**. Aerenchyma formation in plants. In: van Dongen JT, Licausi F, eds. Plant Cell Monographs. Low-Oxygen Stress in Plants. Springer, Vienna, 247–265.

Tamura K, Seki H, Suzuki H, Kojoma M, Saito K, Muranaka T. **2017**. CYP716A179 functions as a triterpene C-28 oxidase in tissue-cultured stolons of *Glycyrrhiza uralensis*. *Plant Cell Reports* **36**: 437–445.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. **2013**. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725–2729.

Tamura K, Yoshida K, Hiraoka Y, Sakaguchi D, Chikugo A, Mochida K, Kojoma

M, Mitsuda N, Saito K, Muranaka T, *et al.* 2018. The basic helix–loop–helix transcription factor GubHLH3 positively regulates soyasaponin biosynthetic genes in *Glycyrrhiza uralensis*. *Plant and Cell Physiology* **59**: 783–796.

Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A. **2014**. Triterpene biosynthesis in plants. *Annual Review of Plant Biology* **65**: 225–257.

Tsuno Y, Fujimatsu T, Endo K, Sugiyama A, Yazaki K. **2018**. Soyasaponins: a new class of root exudates in soybean (*Glycine max*). *Plant and Cell Physiology* **59**: 366–375.

Van Moerkercke A, Steensma P, Gariboldi I, Espoz J, Purnama PC, Schweizer F, Miettinen K, Vanden Bossche R, De Clercq R, Memelink J, *et al.* 2016. The basic helix-loop-helix transcription factor BIS2 is essential for monoterpenoid indole alkaloid production in the medicinal plant *Catharanthus roseus*. *The Plant Journal* **88**: 3–12.

Van Moerkercke A, Steensma P, Schweizer F, Pollier J, Gariboldi I, Payne R, Vanden Bossche R, Miettinen K, Espoz J, Purnama PC, *et al.* 2015. The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. *Proceedings of the National Academy of Sciences* **112**: 8130–8135.

Wang L, Tang W, Hu Y, Zhang Y, Sun J, Guo X, Lu H, Yang Y, Fang C, Niu X, *et al.* 2019. A MYB/bHLH complex regulates tissue-specific anthocyanin biosynthesis in the inner pericarp of red-centered kiwifruit *Actinidia chinensis* cv. Hongyang. *The Plant Journal* **99**: 359–378.

Wu F-H, Shen S-C, Lee L-Y, Lee S-H, Chan M-T, Lin C-S. **2009**. Tape-*Arabidopsis* Sandwich - a simpler *Arabidopsis* protoplast isolation method. *Plant Methods* **5**: 16.

Yamada Y, Kokabu Y, Chaki K, Yoshimoto T, Ohgaki M, Yoshida S, Kato N, Koyama T, Sato F. **2011**. Isoquinoline alkaloid biosynthesis is regulated by a unique bHLH-type transcription factor in *Coptis japonica*. *Plant and Cell Physiology* **52**: 1131–1141.

Yamamoto K, Takahashi K, Caputi L, Mizuno H, Rodriguez-Lopez CE, Iwasaki T, Ishizaki K, Fukaki H, Ohnishi M, Yamazaki M, *et al.* 2019. The complexity of intercellular localisation of alkaloids revealed by single-cell metabolomics. *New Phytologist* **224**: 848–859.

Yamauchi T, Shimamura S, Nakazono M, Mochizuki T. **2013**. Aerenchyma formation in crop species: A review. *Field Crops Research* **152**: 8–16.

Yin J, Li X, Zhan Y, Li Y, Qu Z, Sun L, Wang S, Yang J, Xiao J. **2017**. Cloning and expression of *BpMYC4* and *BpbHLH9* genes and the role of *BpbHLH9* in triterpenoid synthesis in birch. *BMC Plant Biology* **17**: 214.

Yoshida K, Sakamoto S, Kawai T, Kobayashi Y, Sato K, Ichinose Y, Yaoi K, Akiyoshi-Endo M, Sato H, Takamizo T, *et al.* 2013. Engineering the *Oryza sativa* cell wall with rice NAC transcription factors regulating secondary wall formation. *Frontiers in Plant Science* **4**: 383.

Zhang X, Ge F, Deng B, Shah T, Huang Z, Liu D, Chen C. **2017**. Molecular Cloning and Characterization of PnbHLH1 Transcription Factor in *Panax notoginseng*. *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry* **22**: 1268.

903 **Tables**

904 **Table 1. Co-expression analysis of triterpene biosynthetic genes and 23 candidate**

905 **TF genes**

Name	Gene ID		Pearson's correlation coefficient*			
	Gifu B-129 ver 1.2	Miyakojima MG-20 ver 3.0	*OSC3*	*HMGR1*	*CYP716A51*	*SQE1*
OSC3	LotjaGi2g1v0235400	Lj2g3v0914630	1.00	0.51	0.49	0.50
HMGR1	LotjaGi2g1v0352700	Lj2g3v1574250	0.51	1.00	0.66	0.68
CYP716A51	LotjaGi4g1v0438900	Lj4g3v3014670	0.49	0.66	1.00	0.63
SQE1	LotjaGi5g1v0037500	Lj5g3v0347610	0.50	0.68	0.63	1.00
NAC	LotjaGi1g1v0222300	Lj0g3v0171479	0.20	0.24	0.25	0.31
bHLH50 (Ia)	**LotjaGi1g1v0340900**	**Lj1g3v1525930**	**0.57**	**0.68**	**0.58**	**0.85**
MYB	LotjaGi1g1v0548000	-	-	-	-	-
MYB	LotjaGi1g1v0650400	Lj1g3v4012910	-0.54	-0.47	-0.54	-0.75
bHLH (Ib)	LotjaGi1g1v0667300	Lj0g3v0202379	0.23	0.14	0.07	0.48
WRKY	LotjaGi1g1v0694700	Lj1g3v4578840	-0.13	-0.03	-0.21	-0.25
bZIP	**LotjaGi1g1v0724300**	**Lj1g3v4315340**	**0.44**	**0.62**	**0.44**	**0.54**
WRKY	**LotjaGi1g1v0754400**	**Lj1g3v4830200**	**0.63**	**0.77**	**0.54**	**0.64**
Zinc finger	LotjaGi2g1v0209200	Lj2g3v0740620	-	-	-	-
LBD	**LotjaGi2g1v0242300**	**Lj2g3v0934070**	**0.58**	**0.79**	**0.81**	**0.82**
ERF	LotjaGi3g1v0009300	Lj0g3v0243159	-0.28	-0.48	-0.27	-0.39
LBD	LotjaGi3g1v0035700	Lj3g3v0397200	-0.01	0.30	0.67	0.26
SCREAM-like protein	LotjaGi3g1v0203400	Lj3g3v1404080	-	-	-	-
MYB_related	LotjaGi3g1v0380600	Lj0g3v0318769	-0.29	-0.24	-0.07	-0.11
NAC	LotjaGi4g1v0002900	Lj4g3v0050650	-0.49	-0.34	-0.31	-0.55
bHLH32 (IVa2)	**LotjaGi4g1v0043000**	**Lj0g3v0292969**	**0.74**	**0.79**	**0.85**	**0.67**
WRKY	**LotjaGi4g1v0379800**	**Lj4g3v2616130**	**0.67**	**0.70**	**0.47**	**0.69**
ERF	LotjaGi4g1v0422400	Lj4g3v2871880	0.39	0.50	0.34	0.56
Zinc finger	**LotjaGi5g1v0010400**	**Lj5g3v0165540**	**0.66**	**0.66**	**0.56**	**0.82**
NAC	LotjaGi5g1v0158000	Lj5g3v1093960	0.38	0.39	0.25	0.22
Zinc finger	LotjaGi5g1v0173100	-	-	-	-	-
LBD	LotjaGi5g1v0323400	Lj5g3v2099760	0.10	0.09	0.10	0.44
bHLH (IVc)	LotjaGi6g1v0344900	Lj6g3v2168640	-0.38	-0.43	-0.39	-0.52

906 *Pearson's correlation coefficient (PCC) values based on transcriptome data of MG-20 deposited in

907 *Lotus* Base. Bold indicates putative TF genes showing a > 0.4 of PCC to four BA biosynthetic genes.

908 **Table 2. Changes in the expression of triterpene biosynthetic genes**

Pathway	Name	Hy (vs So)	SA (vs SM)	LjbHLH32-OX (vs GFP)	LjbHLH50-OX (vs GFP)
TFs	*bHLH32*	**1.53**	**3.07**	**5.49**	1.00
	bHLH50	**1.68**	**2.01**	**-0.76**	**3.00**
Precursor	*HMGR1*	**1.53**	**4.03**	**2.88**	0.39
	SQE1	**1.22**	**3.02**	**1.30**	-0.11
Betulinic acid	*OSC3 (LUS)*	**4.24**	**2.73**	**4.18**	**2.42**
	CYP716A51	**3.92**	**4.02**	**3.18**	**1.41**
Soyasaponin /Oleanolic acid	*OSC1 (bAS)*	-0.21	**-1.17**	0.038	0.38
Soyasaponin	*LjCYP93E1*	-0.59	-0.59	-0.20	-0.023
	LjCYP72A61	-0.87	0.56	0.20	-0.11
	CSyGT	-0.47	0.28	0.11	0.27
	DDMPT	-0.87	0.79	0.20	-0.21
Ursolic acid	*OSC9 (aAS)*	-	-2.23	-	-
Sterol	*OSC5 (CAS)*	-0.2	-1.09	0.046	0.51
	OSC7 (LAS)	0.57	0.62	-0.30	0.43
AMY2 gene cluster	*AMY2 (mixAS)*	-0.95	-0.032	**1.56**	**-0.82**
	CYP71D353	1.03	-1.15	0.54	0.36
	CYP88D4	**2.14**	**4.34**	**1.43**	0.41
	CYP88D5	**1.13**	**2.25**	**1.48**	0.18

909 Log base 2 values of fold changes of expression levels are listed. Positive values mean upregulation
910 in BA-rich samples relative to control samples. Differentially expressed genes are indicated in bold.
911 Details are provided in Table S7.

912

913

914

915

916 **Figure Legends**

917 **Fig. 1. Triterpene sapogenin profiling by gas chromatography-mass spectrometry**
918 **(GC-MS). (a)** Analysis of *Lotus japonicus* adventitious roots (AR) and whole roots of
919 soil-grown and hydroponically cultured plants. Increased betulinic acid (BA)
920 accumulation was observed in the extracts of hydroponic roots. A BA peak was
921 observed without acid hydrolysis during triterpene extraction. **(b)** Analysis of Lj
922 suspension cultured cells (LjCC) with and without methyl jasmonate (MeJA). Extracted
923 ion chromatograms (*m/z* = 189 and 306) are shown to assess the presence of BA and
924 soyasapogenol B (SB). **(c)** Extracted ion chromatograms (*m/z* = 189) of hydroponically
925 cultured and soil-grown roots and AR of *Medicago truncatula* cv. R108 and Jemalong
926 A17. The following authentic standards were analyzed (Suzuki *et al.*, 2019): β-amyrin
927 (1), α-amyrin (2), lupeol (3), 24-OH-β-amyrin (4), oleanolic acid (5), BA (6),
928 sophoradiol (7), ursolic acid (8), soyasapogenol E (9), SB (10), soyasapogenol A (11).

929 **Fig. 2. Accumulation of betulinic acid (BA) in secondary aerenchyma (SA) tissue**
930 **induced by hydroponic culture. (a)** Images of sections of hypocotyls. White, spongy
931 tissue (SA) had developed on the surface of hypocotyls and roots of *L. japonicus*
932 cultured hydroponically. Quantitation of BA **(b)** and soyasapogenol B **(**SB; **c)**.
933 Hydroponic roots and SA accumulated more BA than soil-grown roots (Soil) and stele,
934 respectively. Means ± standard errors (SEs, n = 3). LR, lateral roots.

935 **Fig. 3. Quantitative reverse transcription-PCR (qRT-PCR) analysis of triterpene**
936 **biosynthetic genes.** Relative expression levels of betulinic acid (BA) and soyasaponin
937 biosynthesis genes were measured in triplicate. The values were normalized to the
938 expression level of a reference gene (*LjUBQ1*). Means ± standard deviations (SDs). Soil,
939 soil-grown roots; hydroponics, hydroponically grown roots; SM, secondary meristem;

940 secondary aerenchyma, SA.

941 **Fig. 4. Selection of candidate transcription factors (TFs) regulating betulinic acid**
942 **(BA) biosynthesis.** Principal component analysis (PCA) of the transcriptomes of
943 soil-grown (So) and hydroponically cultured (Hy) roots **(a)**, and secondary meristem
944 (SM) and secondary aerenchyma (SA) tissues **(b)**. **(c)** Venn diagram of differentially
945 expressed genes (DEGs) upregulated in Hy and SA compared to So and SM,
946 respectively. **(d)** Phylogenetic relationships of basic helix-loop-helix proteins. TFs
947 known to regulate triterpene biosynthesis are highlighted by black circles. Blue and red
948 circles indicate LjbHLH32 and LjbHLH50, respectively. ER, endoplasmic reticulum;
949 OA, oleanolic acid; IQA, isoquinoline alkaloid.

950 **Fig. 5. Metabolite and gene expression analysis of LjbHLH32- and**
951 **LjbHLH50-overexpressing (OX) transgenic hairy root lines. (a)** Triterpenes were
952 extracted from three control (GFP-expressing) lines and a basic helix-loop-helix
953 (bHLH)-OX line. Accumulation of betulinic acid (BA) and soyasapogenol B (SB) in
954 hairy roots was analyzed by gas chromatography-mass spectrometry (GC-MS). Relative
955 amounts were calculated from the peak areas of BA and SB. **(b)** Relative expression
956 levels of bHLH and biosynthetic genes by quantitative reverse transcription-PCR
957 (qRT-PCR). Error bars indicate standard deviations (SD, n = 3).

958 **Fig. 6. RNA-sequencing (RNA-seq) analysis of LjbHLH32- and**
959 **LjbHLH50-overexpressing (OX) transgenic hairy root lines.** Principal component
960 analysis (PCA) of the transcriptomes of control and LjbHLH32-OX roots **(a)**, and
961 control and LjbHLH50-OX roots **(b)**. **(c)** Venn diagram of 256 differentially expressed
962 genes (DEGs) commonly upregulated in hydroponic whole roots (Hy) and secondary
963 aerenchyma (SA), and DEGs upregulated in bHLH-OX hairy root lines compared to

964 control (GFP-expressing) hairy roots.

965 **Fig. 7. Transient effector-reporter analysis of LjbHLH32 and LjbHLH50 for the**
966 **promoters of triterpene biosynthetic genes.** Effector, reporter, and reference plasmids
967 were co-transfected into *Arabidopsis* mesophyll cell protoplasts. Luciferase (LUC)
968 activity levels were measured after overnight incubation. Error bars indicate standard
969 errors (SE, n = 4). Asterisks indicate significant differences between the negative
970 control (NC) and transcription factors (TFs) by Dunnett's test (*P < 0.05, **P < 0.01,
971 ***P < 0.001). PC, positive control; OA, oleanolic acid; BA, betulinic acid; UA, ursolic
972 acid.

973

974 **Supporting Information**

975 **Fig. S1.** Transfer DNA regions of gateway destination vectors for *L. japonicus*
976 transformation.

977 **Fig. S2.** Triterpene biosynthetic pathways in *L. japonicus*.

978 **Fig. S3.** Mass spectra of peaks 6 and 10 in the chromatograms in Fig. 1.

979 **Fig. S4.** Time-course-dependent increase in BA in hydroponic roots of *L. japonicus*.

980 **Fig. S5.** Magnified image of secondary aerenchyma from a hydroponic *L. japonicus*
981 plant.

982 **Fig. S6.** MA plots for differentially expressed gene (DEG) detection.

983 **Fig. S7.** Protein alignments of the basic domains of clade IVa bHLHs.

984 **Fig. S8.** Alignments of AtSCRM and AtSCRM-like proteins.

985 **Fig. S9.** Phylogenetic tree of LBD (LOB) proteins in *M. truncatula* and *L. japonicus*.

986

987 **Table S1.** List of primers.

988 **Table S2.** List of FastQ files obtained.

989 **Table S3.** Promoter sequences isolated in this study.

990 **Table S4.** List of differentially expressed genes (DEGs) identified in this study.

991 **Table S5.** Annotation of 256 differentially expressed genes (DEGs) commonly

992 upregulated in Hy and SA samples.

993 **Table S6.** Annotation of differentially expressed genes (DEGs) in Fig. 6c.

994 **Table S7.** Detailed results related to triterpene biosynthesis genes obtained by

995 TCC-GUI analysis.

996

997 The English in this document has been checked by at least two professional editors,

998 both native speakers of English. For a certificate, please see:

999 http://www.textcheck.com/certificate/3bPF0g

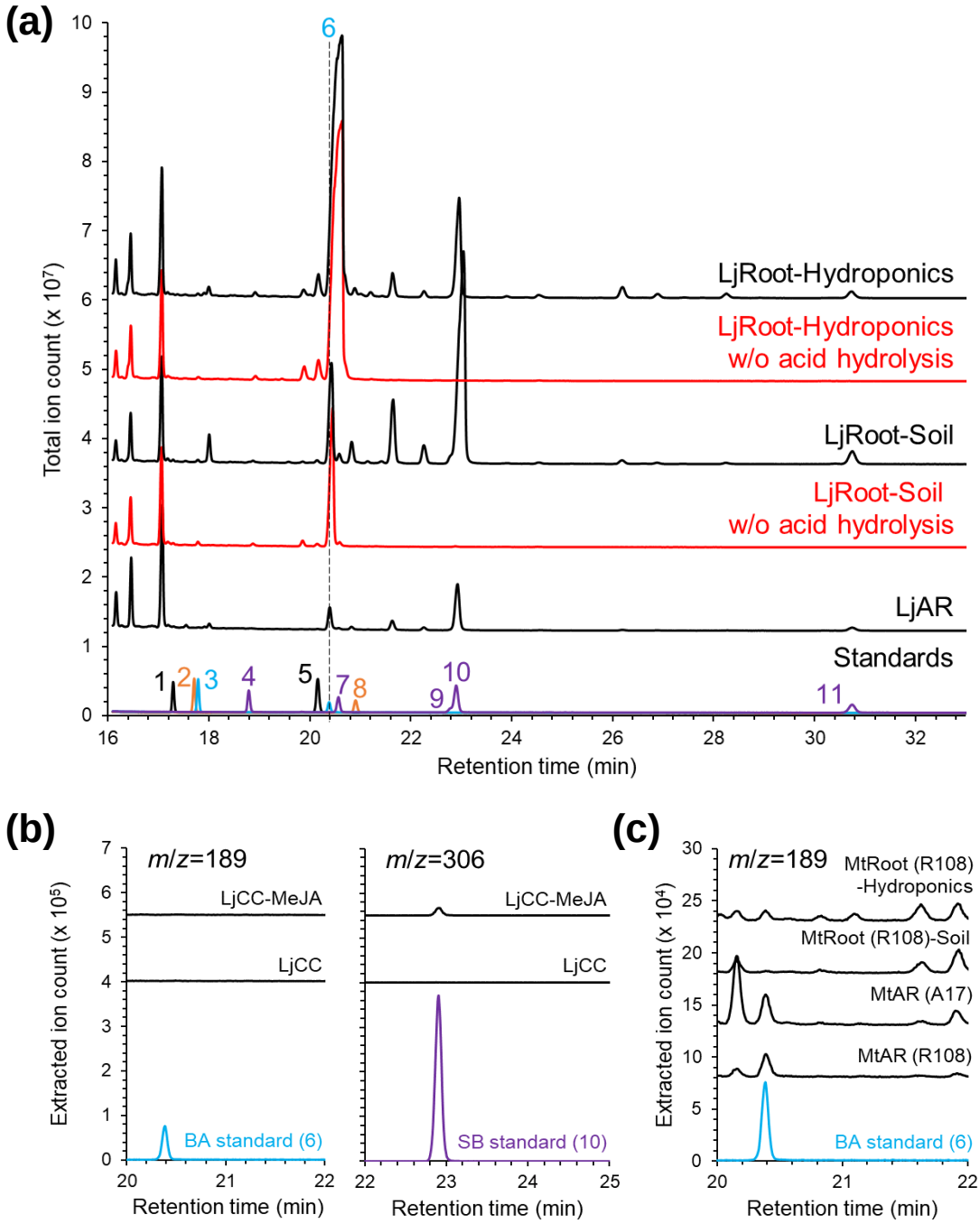
(a)
Total ion count (x 10^7)
Retention time (min)
LjRoot-Hydroponics
LjRoot-Hydroponics w/o acid hydrolysis
LjRoot-Soil
LjRoot-Soil w/o acid hydrolysis
LjAR
Standards
1
2
3
4
5
6
7
8
9
10
11
(b)
m/z=189
m/z=306
LjCC-MeJA
LjCC
BA standard (6)
SB standard (10)
Extracted ion count (x 10^5)
Retention time (min)
(c)
m/z=189
MtRoot (R108) -Hydroponics
MtRoot (R108)-Soil
MtAR (A17)
MtAR (R108)
BA standard (6)
Extracted ion count (x 10^4)
Retention time (min)

Fig. 1

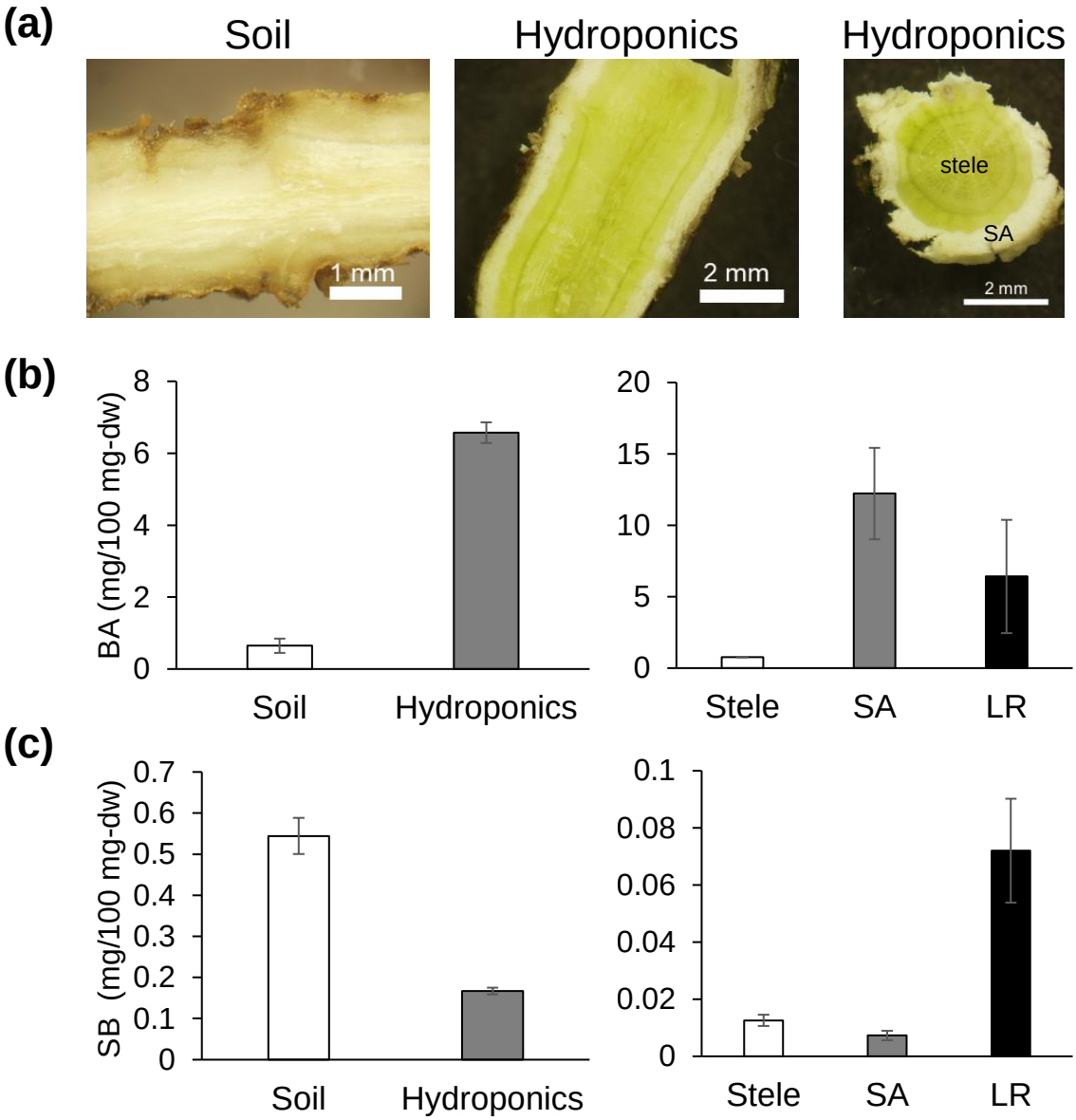
(a)
Soil
Hydroponics
Hydroponics
1 mm
2 mm
stele
SA
2 mm
(b)
BA (mg/100 mg-dw)
8
6
4
2
0
Soil
Hydroponics
20
15
10
5
0
Stele
SA
LR
(c)
SB (mg/100 mg-dw)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Soil
Hydroponics
0.1
0.08
0.06
0.04
0.02
0
Stele
SA
LR

Fig. 2

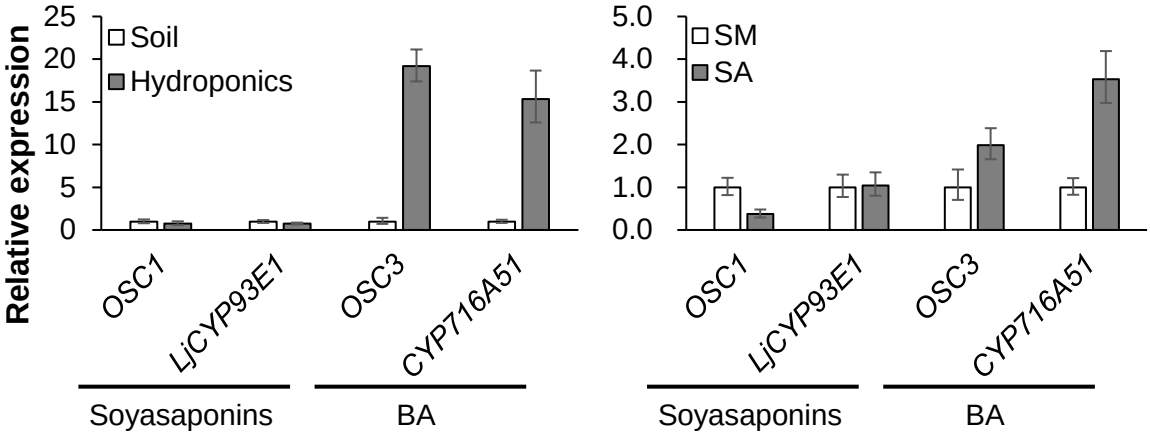
Relative expression
25
20
15
10
5
0
Soil
Hydroponics
OSC1
LjCYP93E1
OSC3
CYP716A51
Soyasaponins
BA
5.0
4.0
3.0
2.0
1.0
0.0
SM
SA
OSC1
LjCYP93E1
OSC3
CYP716A51
Soyasaponins
BA

Fig. 3

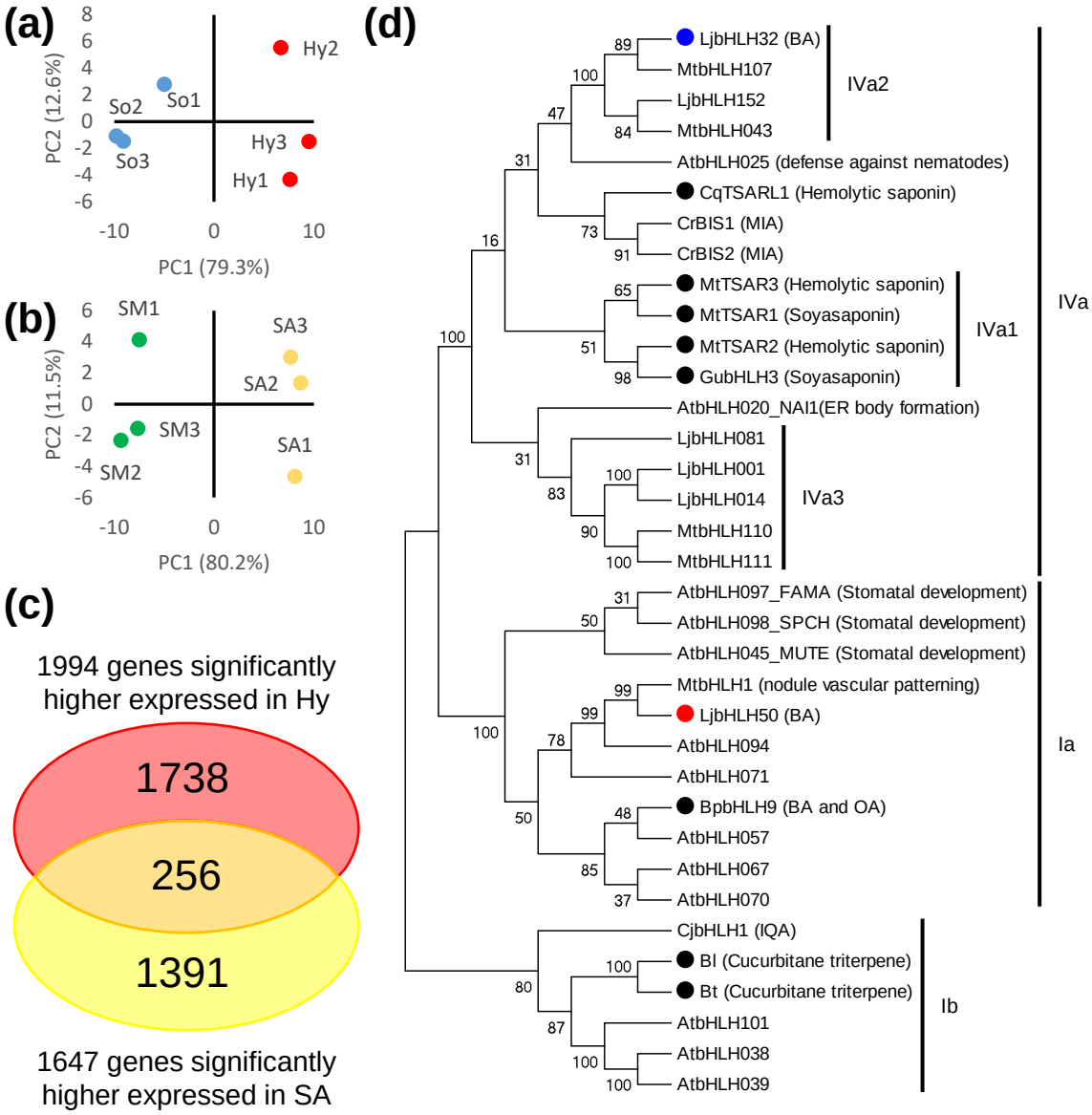
(a)
PC2 (12.6%)
PC1 (79.3%)
Hy2
So1
So2
So3
Hy3
Hy1
(b)
PC2 (11.5%)
PC1 (80.2%)
SM1
SA3
SA2
SM3
SA1
SM2
(c)
1994 genes significantly higher expressed in Hy
1738
256
1391
1647 genes significantly higher expressed in SA
(d)
LjbHLH32 (BA)
MtbHLH107
LjbHLH152
MtbHLH043
IVa2
AtbHLH025 (defense against nematodes)
CqTSARL1 (Hemolytic saponin)
CrBIS1 (MIA)
CrBIS2 (MIA)
MtTSAR3 (Hemolytic saponin)
MtTSAR1 (Soyasaponin)
MtTSAR2 (Hemolytic saponin)
GubHLH3 (Soyasaponin)
IVa1
IVa
AtbHLH020_NAI1(ER body formation)
LjbHLH081
LjbHLH001
LjbHLH014
MtbHLH110
MtbHLH111
IVa3
AtbHLH097_FAMA (Stomatal development)
AtbHLH098_SPCH (Stomatal development)
AtbHLH045_MUTE (Stomatal development)
MtbHLH1 (nodule vascular patterning)
LjbHLH50 (BA)
AtbHLH094
AtbHLH071
BpbHLH9 (BA and OA)
AtbHLH057
AtbHLH067
AtbHLH070
Ia
CjbHLH1 (IQA)
Bl (Cucurbitane triterpene)
Bt (Cucurbitane triterpene)
AtbHLH101
AtbHLH038
AtbHLH039
Ib

Fig. 4

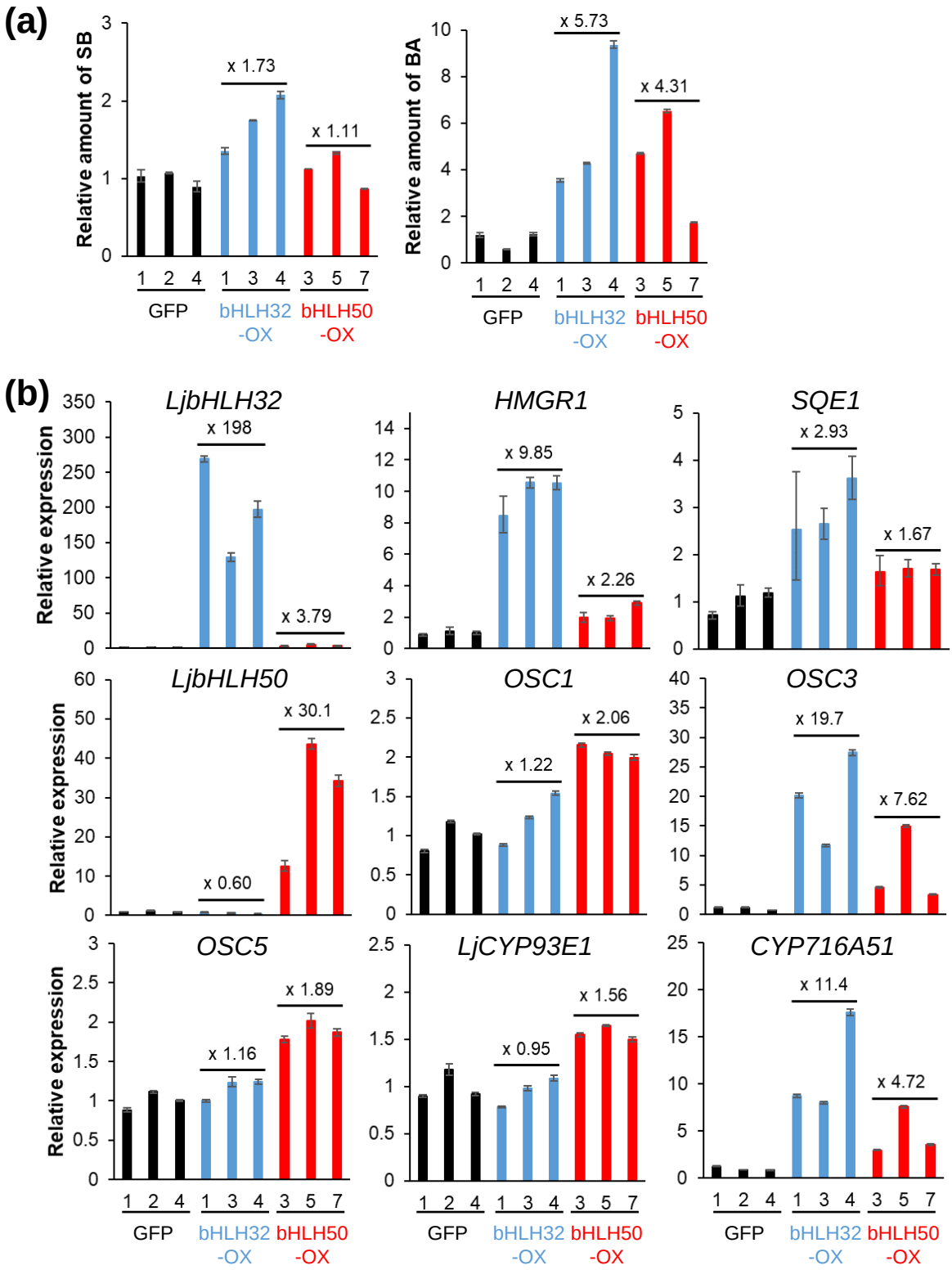
(a)
Relative amount of SB
x 1.73
x 1.11
Relative amount of BA
x 5.73
x 4.31
1 2 4 1 3 4 3 5 7
GFP
bHLH32 -OX
bHLH50 -OX
(b)
Relative expression
LjbHLH32
x 198
x 3.79
HMGR1
x 9.85
x 2.26
SQE1
x 2.93
x 1.67
LjbHLH50
x 30.1
x 0.60
OSC1
x 2.06
x 1.22
OSC3
x 19.7
x 7.62
OSC5
x 1.89
x 1.16
LjCYP93E1
x 1.56
x 0.95
CYP716A51
x 11.4
x 4.72

Fig. 5

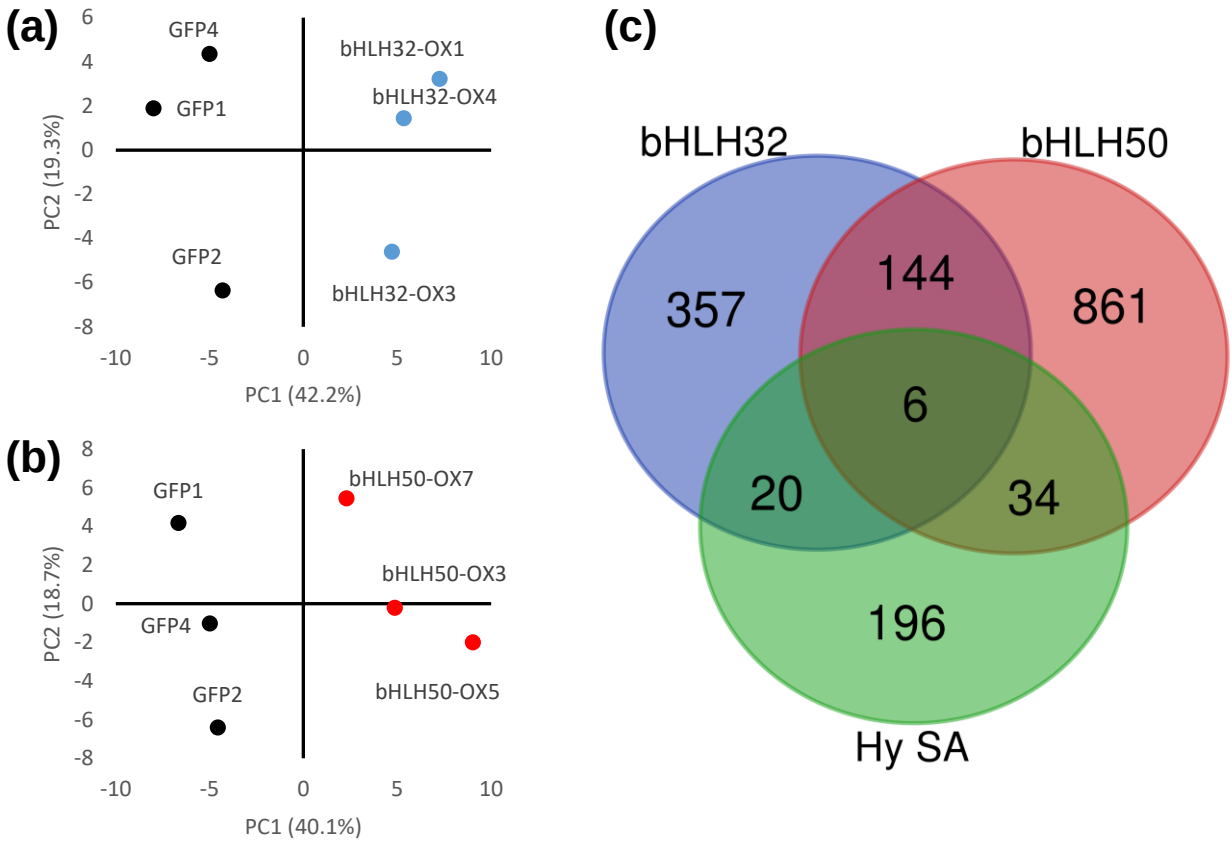
(a)
GFP4
GFP1
GFP2
bHLH32-OX1
bHLH32-OX4
bHLH32-OX3
PC2 (19.3%)
PC1 (42.2%)
(b)
GFP1
GFP4
GFP2
bHLH50-OX7
bHLH50-OX3
bHLH50-OX5
PC2 (18.7%)
PC1 (40.1%)
(c)
bHLH32
bHLH50
Hy SA
357
144
861
6
20
34
196

Fig.6

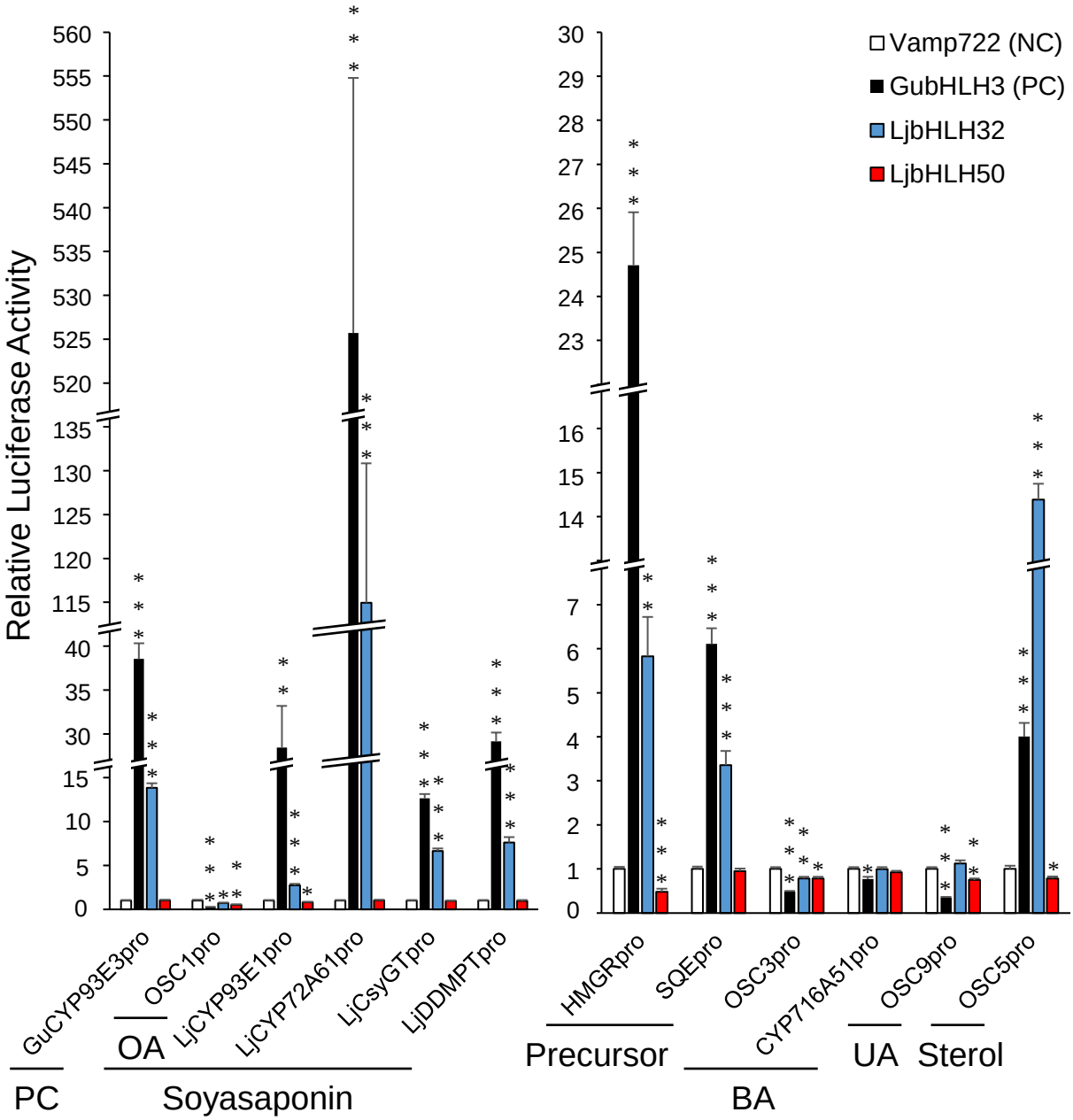
Relative Luciferase Activity
Vamp722 (NC)
GubHLH3 (PC)
LjbHLH32
LjbHLH50
GuCYP93E3pro
OSC1pro
LjCYP93E1pro
LjCYP72A61pro
LjCsyGTpro
LjDDMPTpro
HMGRpro
SQEpro
OSC3pro
CYP716A51pro
OSC9pro
OSC5pro
PC
OA
Soyasaponin
Precursor
BA
UA
Sterol

Fig. 7