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 (4080)

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

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

Running Title: Betulinic acid biosynthesis activators in *L. japonicus*.

33 **Highlight**

34 Hydroponic culture enhanced betulinic acid synthesis in *Lotus japonicus*. RNA-sequencing and
35 functional characterization experiments suggest that LjbHLH32 and LjbHLH50 are the
36 transcription factors activating betulinic acid biosynthesis.

37 **Abstract**

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

55

56 Key words: betulinic acid, bHLH, hydroponics, laser microdissection (LMD), *Lotus*
57 *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

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

113 **Materials and Methods**

114 **Plant materials**

115 Seeds of *L. japonicus* Gifu B-129 accession and *M. truncatula* cv. R108 and Jemalong
116 A17 were sterilized using a five-fold-diluted sodium hypochlorite solution containing
117 0.02% Tween-20 for 15 min with mild agitation. After washing three times in sterilized
118 water, they were soaked in water with mild agitation overnight. The seeds were
119 germinated on 0.8% agar with incubation for 4 days (*L. japonicus*) or 3 days (*M.*
120 *truncatula*) in the dark and for 2 days on a 16 h light/8 h dark cycle at 25°C. The
121 seedlings were transplanted to soil pots or hydroponic solutions and cultured in a plant
122 growth room maintained at 23°C until harvest. We used basal nutrient solution (BNS)
123 for all hydroponic solutions (Conn *et al.*, 2013). *L. japonicus* seedlings were cultured in
124 5 mL sampling tubes for 2 weeks and transferred to 50 mL sampling tubes with BNS. *M.*
125 *truncatula* seedlings were cultured in 50 mL tubes with BNS from transplant to harvest.
126 For induction of adventitious roots, the roots of 1-week-old seedlings were isolated and
127 cultured in 100 mL of half-strength Gamborg's B5 liquid medium supplemented with
128 3% sucrose and 0.2 mg/L of indole-3-butyric acid. The culture was maintained at 23°C
129 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

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

155 We employed a microwave-assisted method for paraffin-embedding using an
156 Energy Beam Sciences H2850 Microwave Tissue Processor (Agawan, MA). A vial
157 containing fixed tissue and pre-chilled methanol was placed on ice and microwaved
158 three times at 250 W at 37°C for 15 min. The methanol was replaced each time. The
159 temperature in the vial was < 20°C during the first and second microwave treatments;
160 however, 37°C was allowed in the third treatment. The solvent was gradually replaced
161 with 1-butanol, using methanol:1-butanol (3:1), methanol:1-butanol (1:1),
162 methanol:1-butanol (1:3), and 1-butanol, with microwaving at each step at 300 W and
163 37°C for 1.5 min. 1-Butanol was gradually replaced with melted paraffin wax
164 (Surgipath Paraplast X-tra Tissue Infiltration/Embedding Media; Leica Biosystems,
165 Wetzlar, Germany) using 1-butanol:paraffin (3:1), 1-butanol:paraffin (1:1), and
166 1-butanol:paraffin (1:3) with microwaving at each step at 300 W and 58°C for 10 min.
167 Two treatments at 250 W at 58°C for 10 min and five treatments at 250 W at 58°C for
168 30 min were performed with new paraffin solution. The specimens were embedded in
169 paraffin wax and cooled to room temperature. The paraffin blocks were stored at 4°C.

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

178 Laser-capture Microdissection System (Zeiss, Jena, Germany) with a MicroBeam Laser
179 (cut energy: 55–58, cut focus: 80, LPC energy: 30, LPC focus: 2). AdhesiveCap 500
180 clear (Zeiss) was used to collect cell layers cut by the laser. RNA was purified using an
181 Arcturus PicoPure RNA Isolation Kit (Thermo Fisher Scientific).
182
183 **Quantitative reverse transcription-PCR (qRT-PCR) analysis**
184 First-strand cDNA was synthesized from purified total RNA using PrimeScript RT
185 Master Mix (Perfect Real Time; Takara Bio, Shiga, Japan) following the manufacturer's
186 instructions. The reaction time was 15 and 60 min, for total RNA of whole roots and
187 that of LMD samples, respectively. We performed qRT-PCR analysis following a
188 previous report (Suzuki *et al.*, 2019) using primer nos. 1–20 in Table S1.
189
190 **RNA-seq**
191 Transcriptome profiling of soil-grown and hydroponic roots was performed with three
192 biological replicates (Table S2). Complementary DNA library construction was
193 performed using a TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA)
194 according to the manufacturer's instructions. The resulting cDNA libraries were
195 sequenced using HiSeq 2500 (Illumina) with 100 bp single-end (SE) reads. Adapter
196 sequences and low-quality reads were removed from raw RNA-seq reads by
197 Trimmomatic v. 0.38 (Bolger *et al.*, 2014). Quantification of transcripts was conducted
198 with Salmon v. 0.14.2 (Patro *et al.*, 2017). Reference *L. japonicus* coding sequences
199 (predicted from genome assembly build v. 1.2 of Gifu B-129 accession; Kamal *et al.*,
200 2020) were downloaded from *Lotus* Base (https://lotus.au.dk/). A mapping-based index
201 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*.

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

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

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

239 The vector backbone, including the gateway cassette and GFP, and fragments,
240 including the *LjUBQ1* promoter, were amplified from pUB-GWS-GFP using primer sets
241 33/36 and 38/39, respectively. These two fragments were assembled with the HSP
242 terminator fragment (primer nos. 34 and 37) to produce pSD9.

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

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

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

249 LR reaction of pSD11 and entry clones (pENTR1A-*GFP*, *LjbHLH32*, and *LjbHLH50*)
250 using Gateway LR Clonase II Enzyme Mix (Thermo Fisher Scientific). These three
251 vectors were introduced into *Agrobacterium rhizogenes* ATCC15834 by electroporation.
252 *Agrobacterium rhizogenes*-mediated hairy root transformation was performed
253 following our previous study (Suzuki *et al.*, 2019), with some modifications. The
254 incubation temperature was maintained at 25°C in the germination, co-culture, and hairy
255 root induction steps. Isolated hairy roots were cultured at 23°C. For liquid culture of
256 hairy roots, the rotary shaker agitation speed was set to 90 rpm. The
257 germination/co-cultivation medium contained 1/2 B5 basal salts. Hairy root elongation
258 (HRE) medium is composed of 1/2 B5 basal salts, 1/4 Gamborg's B5 vitamins, and 300
259 mg/L of cefotaxime sodium. All media were adjusted to pH 5.5 before autoclaving and
260 freshly prepared. Seeds were germinated as described above. After infection and
261 co-cultivation (Suzuki *et al.*, 2019), seedlings were placed on HRE agar supplemented
262 with 1% sucrose and a filter paper strip was placed on the hypocotyls to prevent the
263 hairy roots from drying. Hairy roots were elongated under a 16 h light/8 h dark cycle for
264 3 weeks, followed by shoot removal. Isolated hypocotyls possessing hairy roots were
265 incubated on new HRE agar containing 3% sucrose without light for 10 days. Single
266 hairy roots were isolated and cultured in 5 mL of liquid HRE medium supplemented
267 with 3% sucrose using six-well cell culture plates. They were subcultured twice every
268 16 days. Hairy root lines emitting GFP fluorescence were selected and each was
269 subcultured into six wells. After 26 days of culture, hairy roots were harvested for
270 GC-MS, qRT-PCR, and RNA-seq analysis (150 bp PE reads, Novaseq 6000 platform,
271 by Takara Bio; Table S2) as described above.
272

273 **Cloning of promoter sequences**

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

284

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

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

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

297 (Oshima *et al.*, 2013) to produce three effector constructs (pDEST35SHSP-*GubHLH3*,
298 *LjbHLH32*, and *LjbHLH50*). The plasmid pDEST35SHSP-*vesicle-associated membrane*
299 *protein 722* (*Vamp722*; Yoshida *et al.*, 2013) was used as a negative control.
300 We employed phRLHSP [encoding a modified *Renilla* luciferase (RLUC);
301 Yoshida *et al.*, 2013] as an internal reference to normalize the results of our dual
302 luciferase reporter assays.
303
304 **Transient effector-reporter analysis**
305 *A. thaliana* Col-0 was grown under a 16 h light/8 h dark cycle at 22°C for 3–4 weeks.
306 Rosette leaves were used for protoplast isolation after peeling by the Tape-*Arabidopsis*
307 Sandwich method (Wu *et al.*, 2009). Protoplasts were isolated as described (Yoshida *et*
308 *al.*, 2013) with minor modifications. Peeled leaves were incubated for 1 h in enzyme
309 solution [400 mM mannitol, 20 mM KCl, 10 mM $CaCl_2$, 10 g/L of cellulase onozuka
310 R10 (Yakult Pharmaceutical Industry Co., Ltd., Tokyo, Japan), 2.5 g/L of macerozyme
311 R10 (Yakult Pharmaceutical Industry Co., Ltd.), 10 mM 2-mercaptoethanol, and 20 mM
312 2-morpholinoethanesulfonic acid (MES), pH 5.7] in a shaker (22°C, 60 rpm). The
313 protoplasts were filtered through a 70 μm cell strainer and collected by centrifugation at
314 100 × *g* for 10 min. The supernatant was removed by decantation and the remaining
315 pellet was washed thrice with W5 buffer [150 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl,
316 and 2 mM MES, pH 5.7] followed by centrifugation at 100 × *g* for 5 min. The pellet
317 was washed with MMg solution [400 mM mannitol, 15 mM $MgCl_2$, and 4 mM MES,
318 pH 5.7] and resuspended in 5 mL of MMg solution (when using 24 leaves).
319 Reporter, effector, and reference plasmids were co-transfected into protoplasts
320 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

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

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

364

365 **SA produces and accumulates BA**

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

369 hypocotyl and main roots (Fig. S4A–C). Development of a white, spongy tissue was
370 confirmed by anatomical analysis (Fig. 2A), and we assumed that this tissue was
371 homologous to SA in *G. max* (soybean). Soybean develops SA from the SM on stem,
372 hypocotyl, main and lateral roots, and nodules as an adaptive trait to survive flooding by
373 promoting oxygen diffusion in roots (Shimamura *et al.,* 2010; Yamauchi *et al.,* 2013).
374 Hydroponic culture likely resulted in a lack of oxygen in roots, similar to flooding.
375 Hydroponic roots accumulated 10-fold more BA than soil-grown roots, and its quantity
376 reached 6.5 mg/100 mg-dry weight (Fig. 2B). Isolated SA accumulated a higher amount
377 of BA (> 12 mg/100 mg-dry weight) than stele and lateral roots (Fig. 2B). However, SB
378 did not increase in hydroponic roots and SA (Fig. 2C).

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

388

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

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

393 LMD samples was conducted in three and four replicates, respectively. Data processing
394 was performed in triplicate for all samples as mentioned in Materials and Methods. We
395 performed differential expression analyses independently in data set 1 [soil-grown roots
396 (So) and hydroponic roots (Hy)] (Figs. 4A and S6A) and data set 2 (SM and SA) (Figs.
397 4B and S6B). Based on the MA plot (Fig. S6), we identified 3152 and 1994 DEGs
398 whose expression was higher in So and Hy, respectively (Table S4). SM and SA had
399 2169 and 1647 DEGs with elevated expression levels (Table S4). There were 256 DEGs
400 overlapping between Hy and SA (Fig. 4C), including 23 putative TF genes (Fig. 4C and
401 Table S5), as well as BA biosynthetic genes (*OSC3* and *CYP716A51*) and genes for
402 rate-limiting enzymes for triterpene biosynthesis [3-hydroxy-3-methyl-glutaryl-CoA
403 reductase1 (HMGR1) and squalene epoxidase1 (SQE1)].

404 We assessed the correlation between the expression patterns of BA biosynthetic
405 genes and those of candidate TF genes using the gene expression atlas in *Lotus* Base
406 (https://lotus.au.dk/expat/; 81 conditions based on annotation and prediction in
407 Miyakojima MG-20 genome v. 3.0). Seven TF genes were co-expressed with *OSC3*,
408 *CYP716A51*, *HMGR1*, and *SQE1* (Pearson's correlation coefficient [PCC] threshold:
409 0.4; Table 1). Among the seven TFs, we selected LjbHLH32 (LotjaGi4g1v0043000) and
410 LjbHLH50 (LotjaGi1g1v0340900) for further analysis because bHLH TFs are triterpene
411 biosynthesis regulators in Fabaceae (Mertens *et al.*, 2016; Tamura *et al.*, 2018), *Panax*
412 spp. (Zhang *et al.*, 2017), cucumber (Shang *et al.*, 2014), and *B. platyphylla* (Yin *et al.*,
413 2017).

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

including representative bHLH proteins involved in specialized plant metabolism (Fig. 4D). Fabaceae clade IVa bHLH TFs were further subdivided into three subclades, and subclade IVa1 included all Fabaceae triterpene saponin biosynthesis activators identified in clade IVa (Suzuki *et al.*, 2021). Interestingly, LjbHLH32 was a member of subclade IVa2 but not IVa1 (Fig. 4D). The orthologs in subclade IVa2 were highly conserved in Fabaceae plants and were commonly expressed in roots and nodules of model Fabaceae plants (Suzuki *et al.*, 2021). However, no subclade IVa2 bHLHs have been functionally characterized to date. The closest homolog of another candidate LjbHLH50 in *M. truncatula* is MtbHLH1 (Medtr3g099620), which is suggested to regulate nitrogen transfer between nodule and root cells by controlling nodule vascular patterning (Godiard *et al.*, 2011). LjbHLH50 and MtbHLH1 showed 60% amino acid identity (99% coverage) and belonged to clade Ia (Fig. 4D), similar to BpbHLH9, an OA and BA biosynthesis regulator in *B. platyphylla* (Yin *et al.*, 2017). These phylogenetic relationships between our candidates and known triterpene biosynthesis regulators suggest that LjbHLH32 and LjbHLH50 are involved in BA biosynthesis regulation in *L. japonicus*.

Overexpression of LjbHLH32 or LjbHLH50 enhances BA production in transgenic hairy roots

We generated GFP-expressing (control), LjbHLH32-overexpressing (LjbHLH32-OX), and LjbHLH50-OX hairy root lines for functional identification of LjbHLH32 and LjbHLH50 *in planta*. We chose three independent lines for each construct and analyzed their triterpene content by GC-MS (Fig. 5A). Compared to the control lines, LjbHLH32-OX and LjbHLH50-OX hairy roots accumulated 5.73- and 4.31-fold more

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

450 We performed an RNA-seq analysis of nine hairy root lines (Fig. 5) to test the
451 effect of LjbHLH32 and LjbHLH50 overexpression on a genome-wide scale. The
452 control lines and bHLH-OX lines were distinguished by principal component analysis
453 (PCA) of transcriptome data (Fig. 6A and B). DEGs were identified based on an MA
454 plot (Fig. S6C and D). Compared to the control lines, 527 and 1045 genes were
455 significantly upregulated in LjbHLH32-OX and LjbHLH50-OX, respectively (Table S4).
456 Among the DEGs upregulated in bHLH-OX hairy roots and 256 DEGs commonly
457 upregulated in Hy and SA samples (Fig. 4C), 6 overlapped (Fig. 6C). These six genes
458 included the BA biosynthetic genes *OSC3* and *CYP716A51* but no genes involved in
459 other triterpene biosynthesis pathways (Table S6). Some of the enzyme genes putatively
460 involved in precursor supply for BA biosynthesis were significantly upregulated by
461 overexpression of bHLHs (Table S7), which may contribute to increased BA production
462 in hairy roots (Fig. 5A).

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

465 transport of metabolites, signaling, and defense responses (Table S6). *LjHMGR1* and
466 *LjSQE1* were significantly upregulated by LjbHLH32-overexpression (Table 2).
467 Interestingly, LjbHLH32-OX hairy roots showed significantly elevated expression of
468 genes (*AMY2*, *CYP88D4*, and *CYP88D5*) in the *AMY2* gene cluster (Tables 2 and S7;
469 Krokida *et al.,* 2013). The products of the *AMY2* gene cluster in *L. japonicus* are
470 unknown. AMY2 produced β-amyrin and lupeol in a yeast expression system
471 (Iturbe-Ormaetxe *et al.,* 2003), and β-amyrin and dihydro-lupeol by agro-infiltration
472 into tobacco leaves (Krokida *et al.,* 2013). CYP88D4 and CYP88D5, homologs of
473 CYP88D6 (involved in glycyrrhizin biosynthesis in *G. uralensis*), did not oxidize the
474 tested triterpenes (Seki *et al.,* 2008; Krokida *et al.,* 2013). CYP71D353 showed
475 dihydro-lupeol oxidation activity when transiently expressed in tobacco leaves (Krokida
476 *et al.,* 2013); however, the expression level of *CYP71D353* was not affected by
477 overexpression of LjbHLH32 (Table 2).
478 LjbHLH50-overexpression resulted in significant upregulation of several of the
479 23 TF genes selected by RNA-seq (Table 1), but those TF genes were not upregulated
480 by LjbHLH32 overexpression (Table S6).
481
482 **LjbHLH32 shows similar trends to GubHLH3 in a transient effector-reporter**
483 **analysis**
484 Analyses of hairy roots indicated that overexpression of LjbHLH32 or LjbHLH50
485 upregulated BA biosynthesis directly or indirectly. To obtain insight into how they
486 regulate BA biosynthesis, we examined the transactivation abilities of LjbHLH32 and
487 LjbHLH50 against the promoter sequences of triterpene biosynthetic genes (Table S3)
488 by transient effector-reporter analysis. Effector (*35S*::*bHLH*), reporter

489 (*promoter*::*FLUC*), and reference (*35S*::*RLUC*) constructs were co-transfected into
490 *Arabidopsis* mesophyll cell protoplasts and transactivation activities were measured as
491 relative luciferase activities (FLUC/RLUC; Fig. 7). We used the combination of
492 *GubHLH3* and *GuCYP93E3pro* as the positive control and confirmed a similar level of
493 transactivation (39-fold) as a previous transient effector-reporter analysis using tobacco
494 BY-2 cell protoplasts (Tamura *et al.*, 2018). Similarly, GubHLH3 transactivated
495 promoters of soyasaponin biosynthetic genes in *L. japonicus* except *OSC1pro*
496 (*LjCYP93E3pro*, *LjCYP72A61pro*, *LjCSyGTpro*, and *LjDDMPTpro*) and the genes
497 involved in precursor supply (*LjHMGR1pro* and *LjSQE1pro*). GubHLH3 weakly
498 transactivated the promoter of *OSC5* (*cycloartenol synthase*), consistent with a previous
499 study (Tamura *et al.*, 2018). LjbHLH32 and LjbHLH50 did not transactivate *OSC3pro*
500 and *CYP716A51pro*. Unexpectedly, LjbHLH32 significantly transactivated the
501 promoters of genes in the soyasaponin, precursor, and sterol pathways. Although the
502 fold-changes induced by LjbHLH32 were generally lower than those by GubHLH3, the
503 transactivation pattern of LjbHLH32 was highly correlated to that of GubHLH3
504 (PCC = 0.99). LjbHLH50 did not positively affect any promoters and the results were
505 not correlated to those of GubHLH3 (PCC = 0.36).
506

507 **Discussion**

508 **SA formation and BA biosynthesis**

509 Plant specialized metabolites are often produced and accumulated in specific cell types
510 and organs, such as artemisinin in *Artemisia annua* (Duke & Paul, 1993; Duke *et al.*,
511 1994), nicotine in tobacco (Katoh *et al.*, 2005), glycyrrhizin in *Glycyrrhiza* spp.
512 (Kojoma *et al.*, 2010; Li *et al.*, 2014), and monoterpene indole alkaloid (MIA) in

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

524 Sucrose supply from leaves is required for SA formation (Striker & Colmer,
525 2017; Takahashi *et al.*, 2018). However, it remains elusive whether sucrose functions
526 only as a nutrient during cell differentiation from SM to SA, or also as a signaling
527 molecule (Tognetti *et al.*, 2013). Hydroponically cultured *L. japonicus* will be useful to
528 investigate such molecular mechanisms of SA formation.

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

537

538 **Subclade IVa1 and IVa2 bHLHs are likely functionally diverged in Fabaceae**
539 **triterpene biosynthesis regulation**

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

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

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

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

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

600

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

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

608 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

633 (*Arabidopsis*), suggesting that these bHLHs interact with unknown TFs or indirectly
634 regulate BA biosynthesis via downstream TFs. The 23 putative TFs found in the first
635 screening are promising targets for further analysis (Table 1). This study is a first step
636 toward understanding the complicated transcriptional regulatory mechanisms of BA
637 biosynthesis in *L. japonicus*.
638
639 **Supplementary data**
640 **Fig. S1.** Transfer DNA regions of gateway destination vectors for *L. japonicus*
641 transformation.
642 **Fig. S2.** Triterpene biosynthetic pathways in *L. japonicus*.
643 **Fig. S3.** Mass spectra of peaks 6 and 10 in the chromatograms in Fig. 1.
644 **Fig. S4.** Time-course-dependent increase in BA in hydroponic roots of *L. japonicus*.
645 **Fig. S5.** Magnified image of secondary aerenchyma from a hydroponic *L. japonicus*
646 plant.
647 **Fig. S6.** MA plots for differentially expressed gene (DEG) detection.
648 **Fig. S7.** Protein alignments of the basic domains of clade IVa bHLHs.
649 **Fig. S8.** Alignments of AtSCRM and AtSCRM-like proteins.
650 **Fig. S9.** Phylogenetic tree of LBD (LOB) proteins in *M. truncatula* and *L. japonicus*.
651
652 **Table S1.** List of primers.
653 **Table S2.** List of FastQ files obtained.
654 **Table S3.** Promoter sequences isolated in this study.
655 **Table S4.** List of differentially expressed genes (DEGs) identified in this study.

656 **Table S5.** Annotation of 256 differentially expressed genes (DEGs) commonly
657 upregulated in Hy and SA samples.
658 **Table S6.** Annotation of differentially expressed genes (DEGs) in Fig. 6C.
659 **Table S7.** Detailed results related to triterpene biosynthesis genes obtained by
660 TCC-GUI analysis.
661

662 **Acknowledgements**
663 We thank Miyazaki University and National BioResource Project (NBRP) for providing
664 pUB-GWS-GFP plasmid and *L. japonicus* Gifu B-129 seeds, RIKEN BRC for
665 providing Lj suspension cell line (rpc00032) through NBRP, the Institut National de la
666 Recherche Agronomique for providing *M. truncatula R108* and Jemalong A17 seeds,
667 Prof. Shuhei Yasumoto (Osaka University) for providing genomic DNA of *A. thaliana*
668 Col-0, and Dr. Atsushi Toyoda (National Institute of Genetics) for construction and
669 sequencing of RNA-seq library of soil-grown and hydroponically-cultured whole roots.
670 We also thank Dr. Nobutaka Mitsuda, Dr. Shingo Sakamoto and Ms. Tomoko Niki
671 (AIST) for technical support on transient effector-reporter analysis.
672

673 **Author Contribution**
674 EOF, TM and HSe conceived and supervised the study; All authors designed the
675 experiments; HSu and HT performed the experiments; HSu wrote the manuscript; HT,
676 EOF, MN, TM and HSe made manuscript revisions. All authors read and approved the
677 final manuscript.
678

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.

Funding

This work was supported by the Ministry of Education, Culture, Sports, Science and Technology (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].

Data Availability

The nucleotide sequence reported in this paper has been submitted to the DNA Data Bank of Japan (DDBJ) under the accession numbers of LC609881 (*LjbHLH32*) and LC609882 (*LjbHLH50*). The raw RNA-seq reads obtained were deposited in the DDBJ Sequence Read Archive (DRA) under the accession number DRA011635.

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, *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, *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, *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, *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, *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, *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, *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, *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, *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, *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, *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, *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.

920 **Tables**

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

922 **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

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

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

925 **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

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

929

930

931

932

933 **Figure Legends**

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

946 **Fig. 2. Accumulation of betulinic acid (BA) in secondary aerenchyma (SA) tissue**
947 **induced by hydroponic culture. (A)** Images of sections of hypocotyls. White, spongy
948 tissue (SA) had developed on the surface of hypocotyls and roots of *L. japonicus*
949 cultured hydroponically. Quantitation of BA **(B)** and soyasapogenol B **(**SB; **C)**.
950 Hydroponic roots and SA accumulated more BA than soil-grown roots (Soil) and stele,
951 respectively. Means ± standard errors (SEs, n = 3). LR, lateral roots.

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

957 secondary aerenchyma, SA.

958 **Fig. 4. Selection of candidate transcription factors (TFs) regulating betulinic acid**
959 **(BA) biosynthesis.** Principal component analysis (PCA) of the transcriptomes of
960 soil-grown (So) and hydroponically cultured (Hy) roots **(B)**, and secondary meristem
961 (SM) and secondary aerenchyma (SA) tissues **(B)**. **(C)** Venn diagram of differentially
962 expressed genes (DEGs) upregulated in Hy and SA compared to So and SM,
963 respectively. **(D)** Phylogenetic relationships of basic helix-loop-helix proteins. TFs
964 known to regulate triterpene biosynthesis are highlighted by black circles. Blue and red
965 circles indicate LjbHLH32 and LjbHLH50, respectively. ER, endoplasmic reticulum;
966 OA, oleanolic acid; IQA, isoquinoline alkaloid.

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

975 **Fig. 6. RNA-sequencing (RNA-seq) analysis of LjbHLH32- and**
976 **LjbHLH50-overexpressing (OX) transgenic hairy root lines.** Principal component
977 analysis (PCA) of the transcriptomes of control and LjbHLH32-OX roots **(A)**, and
978 control and LjbHLH50-OX roots **(B)**. **(C)** Venn diagram of 256 differentially expressed
979 genes (DEGs) commonly upregulated in hydroponic whole roots (Hy) and secondary
980 aerenchyma (SA), and DEGs upregulated in bHLH-OX hairy root lines compared to

981 control (GFP-expressing) hairy roots.

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

990

991

992 The English in this document has been checked by at least two professional editors,
993 both native speakers of English. For a certificate, please see:
994 http://www.textcheck.com/certificate/3bPF0g

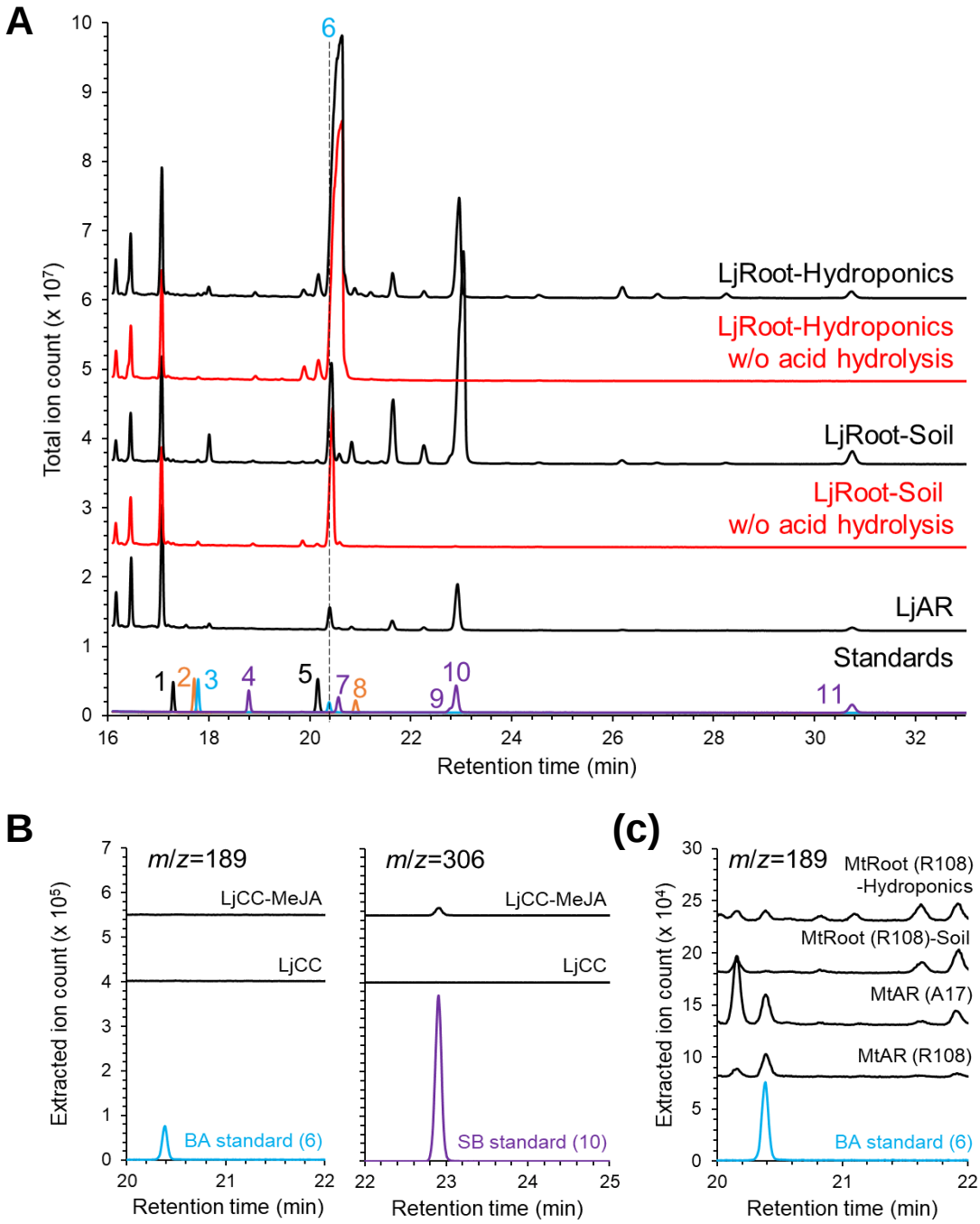

Fig. 1. Triterpene sapogenin profiling by gas chromatography-mass spectrometry (GC-MS). **(a)** Analysis of *Lotus japonicus* adventitious roots (AR) and whole roots of soil-grown and hydroponically cultured plants. Increased betulinic acid (BA) accumulation was observed in the extracts of hydroponic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

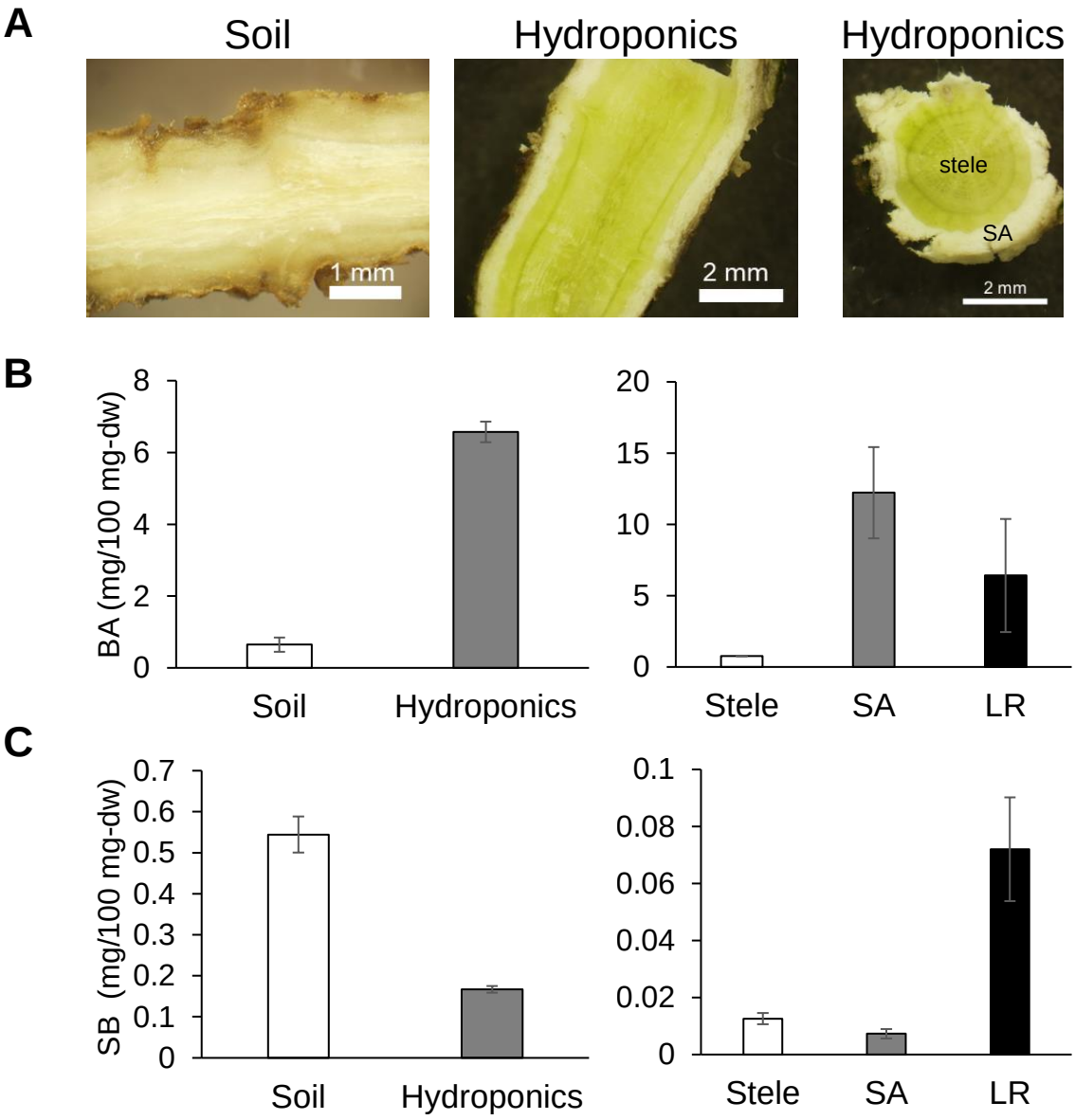

Fig. 2. Accumulation of betulinic acid (BA) in secondary aerenchyma (SA) tissue induced by hydroponic culture. (A) Images of sections of hypocotyls. White, spongy tissue (SA) had developed on the surface of hypocotyls and roots of *L. japonicus* cultured hydroponically. Quantitation of BA **(B)** and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

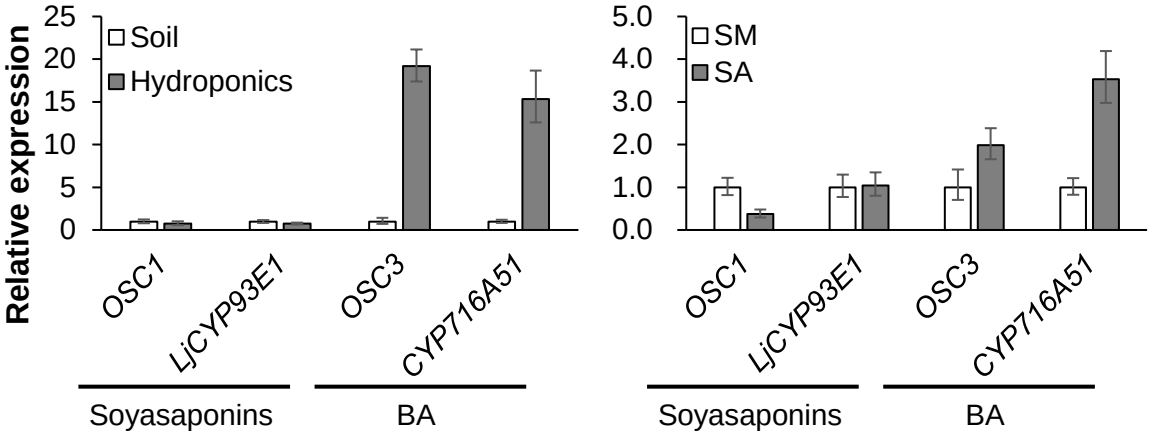

Fig. 3. Quantitative reverse transcription-PCR (qRT-PCR) analysis of triterpene biosynthetic genes. Relative expression levels of betulinic acid (BA) and soyasaponin biosynthesis genes were measured in triplicate. The values were normalized to the expression level of a reference gene (*LjUBQ1*). Means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

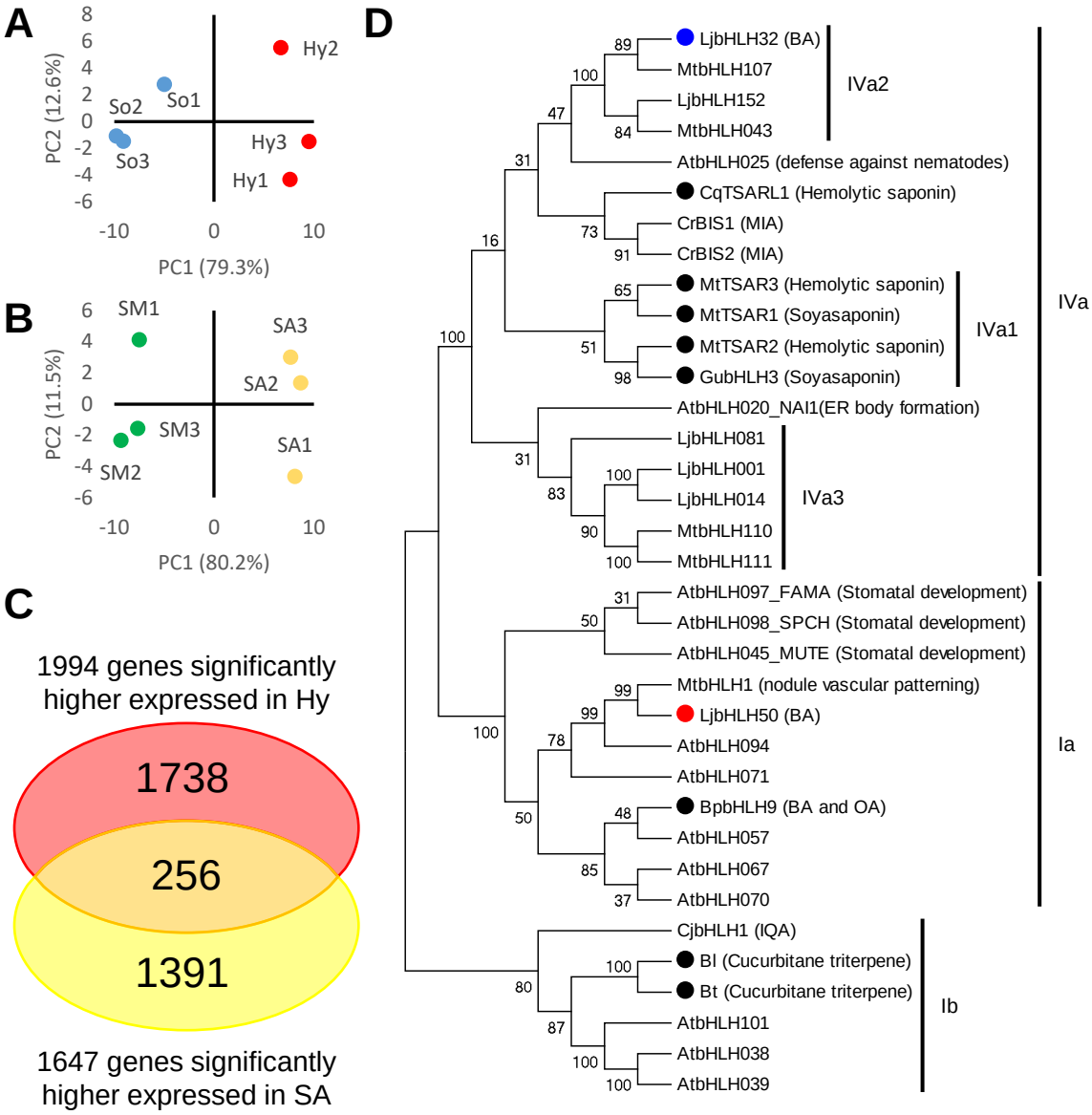

Fig. 4. Selection of candidate transcription factors (TFs) regulating betulinic acid (BA) biosynthesis. Principal component analysis (PCA) of the transcriptomes of soil-grown (So) and hydroponically cultured (Hy) roots **(B)**, and secondary meristem (SM) and secondary aerenchyma (SA) tissues **(B)**.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

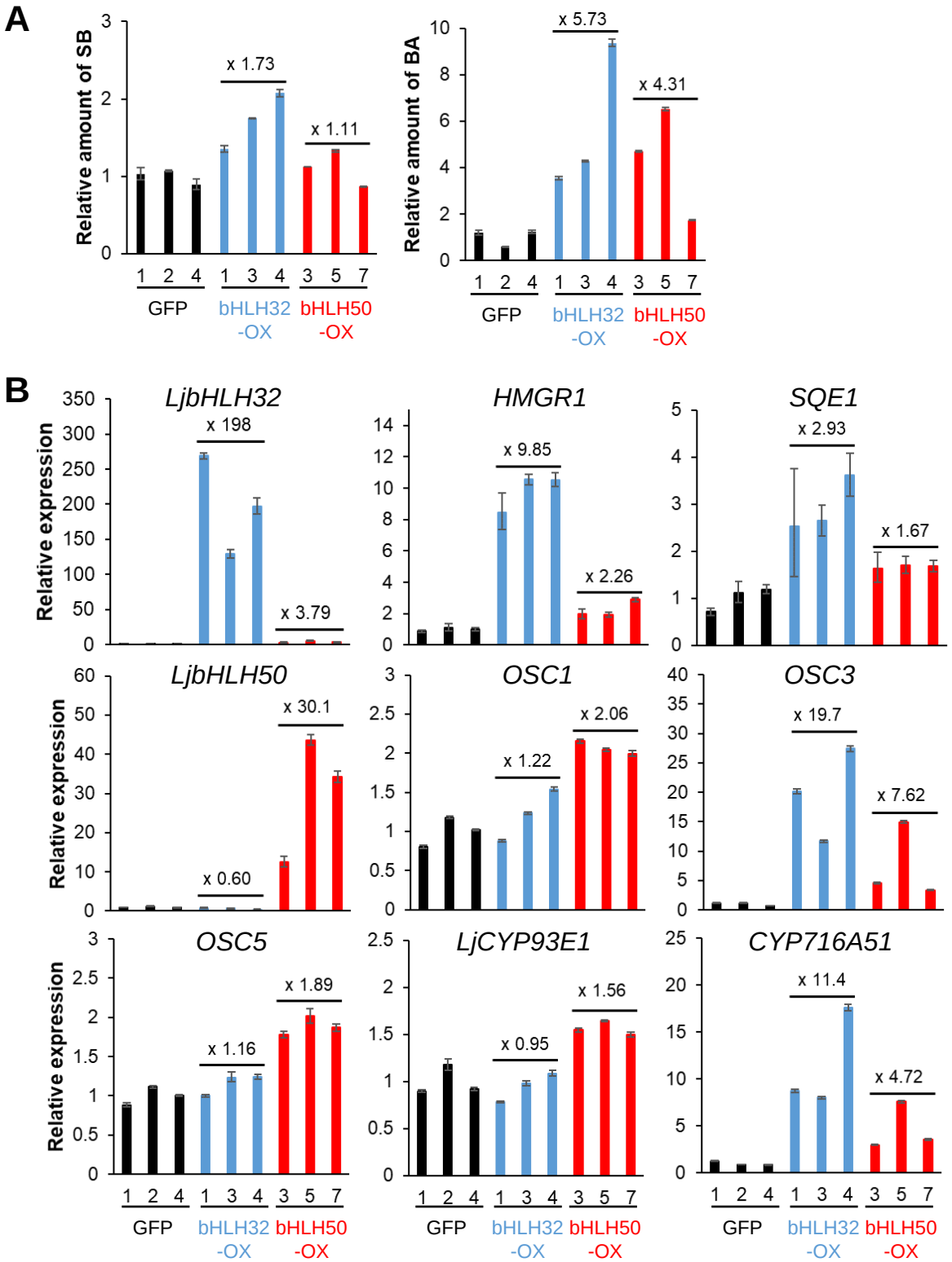

Fig. 5. Metabolite and gene expression analysis of LjbHLH32- and LjbHLH50-overexpressing (OX) transgenic hairy root lines. (A) Triterpenes were extracted from three control (GFP-expressing) lines and a basic helix-loop-helix (bHLH)-OX line. Accumulation of betulinic acid (BA) and soyasapogenol B (SB)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

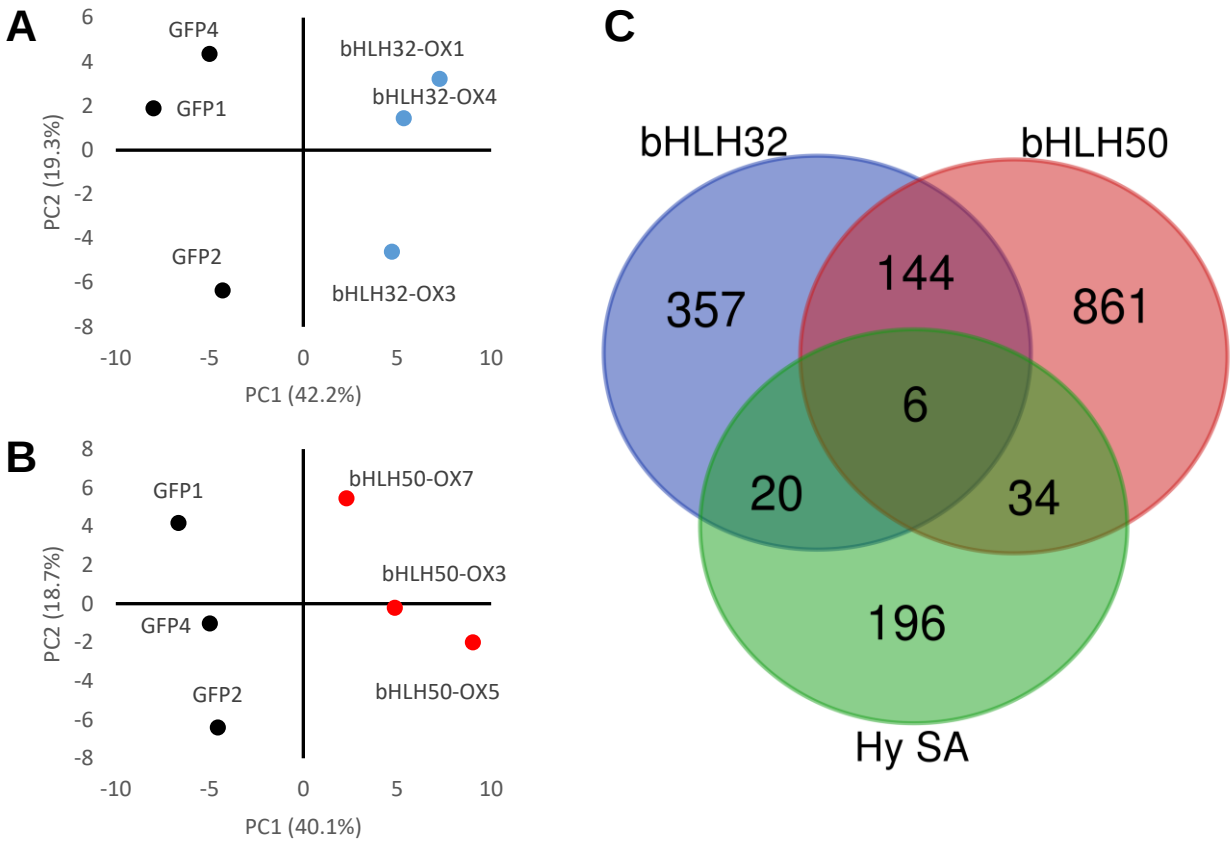

Fig. 6. RNA-sequencing (RNA-seq) analysis of LjbHLH32- and LjbHLH50-overexpressing (OX) transgenic hairy root lines. Principal component analysis (PCA) of the transcriptomes of control and LjbHLH32-OX roots **(A)**, and control and LjbHLH50-OX roots **(B)**. **(C)** Venn diagram of 256 differentially expressed genes (DEGs) commonly upregulated in hydroponic whole roots (Hy) and secondary aerenchyma (SA), and DEGs upregulated in bHLH-OX hairy root lines compared to control (GFP-expressing) hairy roots.

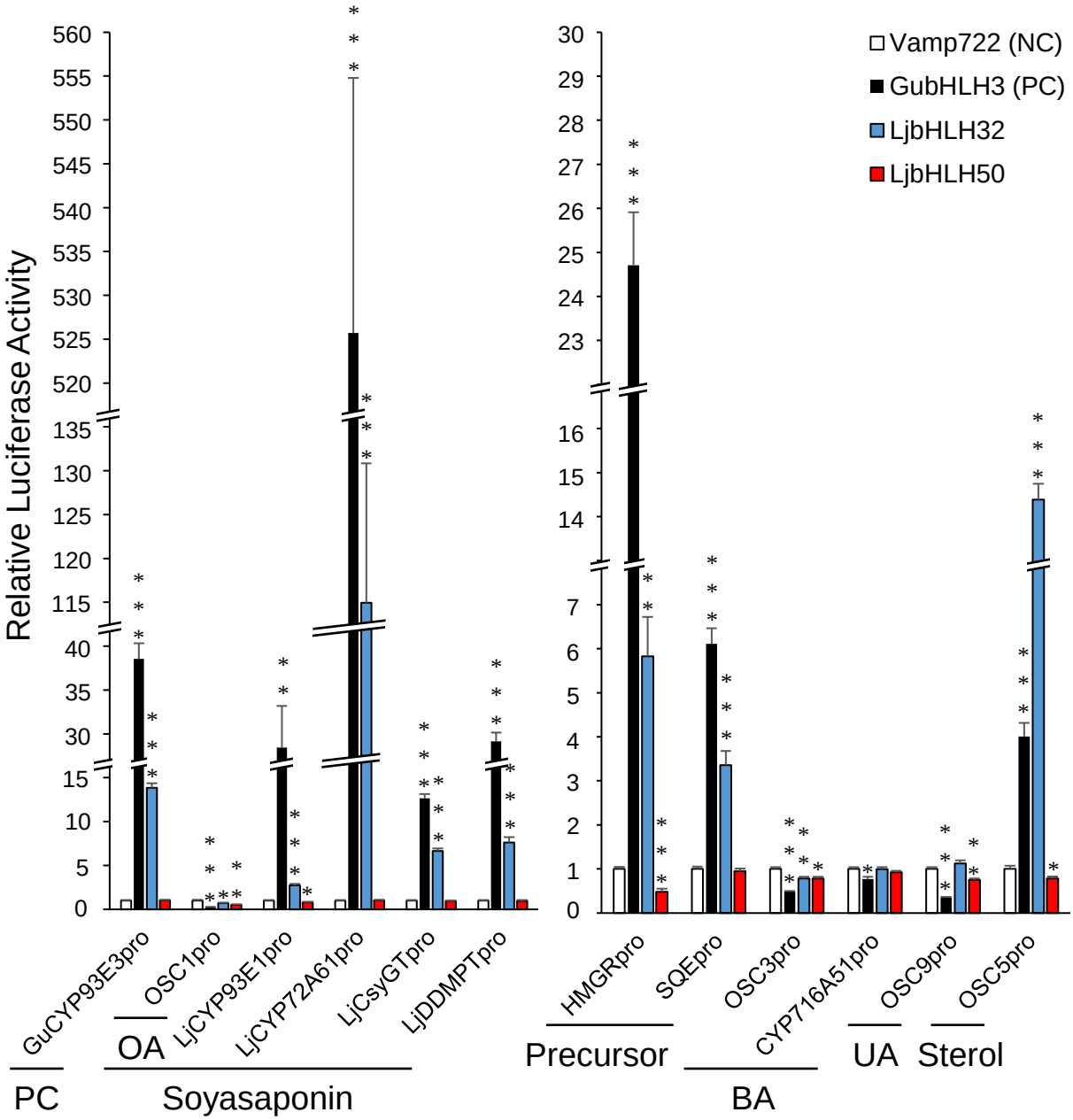

Fig. 7. Transient effector-reporter analysis of LjbHLH32 and LjbHLH50 for the promoters of triterpene biosynthetic genes. Effector, reporter, and reference plasmids were co-transfected into *Arabidopsis* mesophyll cell protoplasts. Luciferase (LUC) activity levels were measured after overnight incubation.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