

Title

Characterization of two G-type half-size ABC transporter genes from the lipid-secreting medicinal plant *Lithospermum erythrorhizon*

The names of the authors

Takuji Ichino^{1, 2}, Kanade Tatsumi^{1, 3}, Yuka Munakata¹, Ai Tsuboyama¹, Eiko Moriyoshi¹, Masaru Nakayasu¹, Kojiro Takanashi⁴, Koichiro Shimomura⁵, Kazufumi Yazaki^{1, *}

The affiliations of the authors

¹Research Institute for Sustainable Humanosphere, Kyoto University, Uji, Japan

²Laboratory of Medicinal Cell Biology, Kobe Pharmaceutical University, Kobe, Japan

³Institute of Plant Molecular Biology, Centre National de la Recherche Scientifique, University of Strasbourg, Strasbourg, France

⁴Department of Biology, Faculty of Science, Shinshu University, Matsumoto, Japan

⁵Graduate School of Life Sciences, Toyo University, Itakura, Japan

*Corresponding Author: Kazufumi Yazaki (yazaki@rish.kyoto-u.ac.jp)

ORCID of the authors

Takuji Ichino, 0000-0001-9058-5660

Kanade Tatsumi, 0000-0002-9810-9367

Eiko Moriyoshi, 0000-0002-8339-3529

Masaru Nakayasu, 0000-0002-6980-7238

Kojiro Takanashi, 0000-0002-4867-5619

Kazufumi Yazaki, 0000-0003-2523-6418

Corresponding author:

Kazufumi Yazaki

Laboratory of Plant Gene Expression, Research Institute for Sustainable Humanosphere, Kyoto

University, Gokasho, Uji 611-0011, Japan

Phone: +81-774-38-4504

Fax: +81-774-38-3600

E-mail: yazaki@rish.kyoto-u.ac.jp

Subject area

(5) Physiology/Biochemistry/Molecular and Cellular Biology

Molecular and Cellular Biology

Number of tables and figures

Number of black and white figures: 0

Number of color figures: 5

Number of tables: 0

Type and number of supplementary materials: 8 supplementary figures and 7 supplementary tables

Abstract

Terrestrial plants can secrete various lipophilic metabolites to facilitate environmental adaptation. Transporters belonging to the G subfamily of ATP-binding cassette (ABCG) proteins are known to be involved in the secretion of lipophilic metabolites from plant cells. The medicinal plant *Lithospermum erythrorhizon* secretes lipophilic specialized metabolites, shikonin derivatives, and echinofurans, as well as triacylglycerol, a major storage lipid. However, the underlying molecular mechanism responsible remains largely uncharacterized. In this study, we identified two genes encoding ABCG transporter proteins from *L. erythrorhizon*, *LeABCG1* and *LeABCG2*, by transcriptome analysis. Phylogenetically, *LeABCG1* is a class A ABCG containing the members that are involved in suberin and exine deposition and jasmonic acid signaling, whereas *LeABCG2* is a class D ABCG involved in cuticle formation. The genomic structures of *LeABCG1* and *LeABCG2* are conserved among the orthologs identified in *Arabidopsis* and tomato. *LeABCG1* is mainly expressed in above-ground tissues, and its expression in root bark is higher than that in peeled roots. In contrast, *LeABCG2* did not show organ-specific gene expression patterns. Exogenous treatment of hairy roots with intracellular and extracellular metabolites of *L. erythrorhizon* slightly affected the expression of these genes. Both the *LeABCG1* and *LeABCG2* proteins localized to the plasma membrane in plant cells, while *LeABCG2* was also distributed in punctate structures on the plasma membrane. Furthermore, *LeABCG1* and *LeABCG2* are the first two ABCG transporter proteins that are actually expressed in *L. erythrorhizon*, suggesting that they are involved in the secretion of lipophilic metabolites across the plasma membrane.

Keywords

ABCG transporter, Lipophilic metabolites, *Lithospermum erythrorhizon*, Plasma membrane, Secretion

Introduction

Plants can protect themselves from biotic and abiotic stresses by secreting lipophilic metabolites to the cell surface, which is the first line of contact with stressors. All land plants commonly transport lipophilic polymers composed of aliphatic precursors and phenolic compounds to the surfaces of specific cells and tissues. Examples of this include the Casparian strip in the root, and the pollen wall, both of which prevent water desiccation, limit UV radiation damage, and control mineral movement. Extracellular hydrophobic biopolymers are classified as waxes, cutins, suberins, and sporopollenins, depending on the site of accumulation and the structural characteristics of the polymer (Franke et al. 2007; Li et al. 2019; Nawrath et al. 2013; Pollard et al. 2008; Samuels et al. 2008). Some plant species also produce small lipophilic specialized metabolites with biological activities related to the stresses of different adaptive niches (Pichersky and Lewinsohn 2011). For example, cucurbitacin, a tetracyclic triterpene in Cucurbitaceae, which is secreted into the soil to modulate the rhizosphere microbiome (Zhong et al. 2022; Zhou et al. 2016). In addition, some volatile organic compounds (VOCs) emitted into the atmosphere also have lipophilic properties (Yazaki et al. 2017).

Deposition of lipophilic metabolites outside plant cells requires transport functions to move the metabolites across the plasma membrane (PM) against the substrate concentration gradient. Recent studies have implicated ATP-binding cassette (ABC) transporter proteins, particularly members of the G subfamily ABCG, in the export of lipophilic compounds into the apoplastic space (Ichino and Yazaki 2022). Membrane-bound ABC transporters consist of a unit, a nucleotide-binding domain (NBD), and a transmembrane domain (TMD), the latter generally consisting of six transmembrane α -helices (Kang et al. 2011). ABC proteins structured as two units in tandem or as a single unit are referred to as full-size ABC transporters and half-size ABC transporters, respectively (Kang et al. 2011).

The first report of an ABCG transporter involved in apoplastic polymer formation in plants is the half-size transporter AtABCG12, which is associated with cuticle formation in *Arabidopsis thaliana* (Pighin et al. 2004). Later, a closely related *A. thaliana* member, AtABCG11, was also reported to be involved in cuticle formation (Bird et al. 2007; McFarlane et al. 2010; Panikashvili et al. 2007). The *Arabidopsis* full-size transporter AtABCG32 and its tomato (*Solanum lycopersicum*) ortholog SLABCG42 have been identified to recognize cutin precursors such as 10,16-dihydroxyhexadecanoyl-2-glycerol, 16-monohydroxyhexadecanoic acid, and hexadecanedioic acid (Elejalde-Palmetto et al. 2021). For suberin, ABCGs such as AtABCG1, AtABCG2, AtABCG6, and AtABCG20 were identified as required for the synthesis of an effective suberin barrier in *A. thaliana* roots and seed coats (Yadav et al. 2014). In the *Arabidopsis* pollen coat, AtABCG26 is thought to translocate sporopollenin precursors (Quilichini et al. 2014). In addition, ABCG transporters are also involved in the excretion of some specialized metabolites, including antifungal terpenoids, and alkaloids. For example, *Petunia x hybrida* PhABCG1 exports major VOCs including methylbenzoate and benzyl alcohol (Adebesin et al. 2017), tobacco *Nb-ABCG1* and *Nb-ABCG2* are putative transporters for an

antimicrobial sesquiterpene capsidiol (Shibata et al. 2016), and camalexin a major phytoalexin in *Arabidopsis* is secreted to the leaf surface by AtABCG36 and AtABCG40 (He et al. 2019).

Lithospermum erythrorhizon (Boraginaceae) is a medicinal herbal plant native to East Asia. This boraginaceous plant produces lipophilic shikonin derivatives (red pigments) that exhibit antibacterial, antioxidant, anti-inflammatory, hemostatic, and antitumor activities (Assimopoulou et al. 2004; Brigham et al. 1999; Yadav et al. 2022; Yazaki 2017). The roots of this plant have also been used as a purple dye to dye cloths for over two thousand years (Ito et al. 2023).

In addition to shikonin derivatives, *L. erythrorhizon* also produces benzoquinone derivatives (e.g., shikono/echinofurans), which are secreted from the cells, while *p*-hydroxybenzoic acid (PHB)-*O*-glucosides, rosmarinic acid, and lithospermic acid B, which accumulate in the cells (Fukui et al. 1984; Yamamoto et al. 2020; Yazaki 2017). Shikonin derivatives accumulate as red granules outside the PM and are often present on cell walls (Kiyoto et al. 2022; Tatsumi et al. 2016, 2023). Thus, the presence of extracellular red granules suggests that novel transport machinery is required for shikonin derivative accumulation. Later biosynthetic reactions for shikonin have been reported to occur in the endoplasmic reticulum (ER) (Ohara et al. 2009; Tatsumi et al. 2023; Wang et al. 2019; Yazaki et al. 2002), suggesting that the intracellular intermediates and/or terminal compounds are transferred to the apoplastic spaces across the PM. In addition, both cultured *L. erythrorhizon* cells and hairy roots secrete significant amounts of triacylglycerol into the apoplast as a matrix lipid to facilitate the secretion of shikonin derivatives (Tatsumi et al. 2023). However, the components responsible for the transport of these lipophilic compounds out of the cell remain poorly understood.

Here, we identified two half-size ABCG transporters, LeABCG1 and LeABCG2, and analyzed their phylogenetic relationships and gene expression patterns, suggesting that both ABCG proteins transport precursors of apoplastic biopolymers and low molecular weight metabolites, including shikonin derivatives and echinofurans. We also showed that these two ABCG proteins function at the PM. Our study provides valuable new information on the transport machinery for specialized metabolites of medicinally useful borage plants.

Materials and methods

Reagents

Phytohormones, indoleacetic acid and kinetin, were purchased from Nacalai Tesque (Kyoto, Japan) and Sigma (St. Louis, MO, USA), respectively. Antibiotics, meropenem trihydrate, spectinomycin dihydrochloride pentahydrate (No. 195-11531), hygromycin B (No. 085-06153), and streptomycin sulfate (No. 190-14342), were also purchased from Wako, Osaka, Japan. Acetylshikonin (NH0602, Nagara Science, Gifu, Japan), 1-hexacosanol (H0342, Tokyo Chemical Industry, Tokyo, Japan), 16-hydroxyhexadecanoic acid (H0675, Tokyo Chemical Industry), 2-monopalmitin (MG (0:0/16:0/0:0), 31-1602-8, Larodan, Solna, Sweden), and 1,2-palmitin-3-palmitolein (TG (16:0/16:0/16:1), 34-1625,

Larodan) were used for chemical treatment. β -Estradiol was also purchased from Wako (No. 056-04044).

Plant materials and growth conditions

Cultured *Lithospermum erythrorhizon* cells of the strain T-TOM were maintained on Linsmaier and Skoog (LS) agar medium (Linsmaier and Skoog 1965) at pH 6.5 containing 3% (w/v) sucrose, 10^{-6} M indoleacetic acid, and 10^{-5} M kinetin at 25 °C in the dark, with subculturing at two-month intervals. Cells were also maintained in LS liquid medium containing the same supplements at the same concentration at 25 °C in the dark, where they were agitated at 80 rpm and cultured at two-week intervals. To induce shikonin production, cells were transferred into M9 liquid medium (Fujita et al. 1981) containing the same concentrations of sucrose and phytohormones and cultured at 25 °C at 80 rpm in the dark for 14 days. Cultured *L. erythrorhizon* cells (strain Mpr) were also used in this study for metabolite extraction (Kiyoto et al. 2022; Nakanishi et al. 2024; Yamamura et al. 2003). Cells were maintained in LS liquid medium at pH 5.7 containing 3% (w/v) sucrose, 10^{-6} M indoleacetic acid, and 10^{-6} M kinetin at 23 °C with shaking at 100 rpm in the dark and subcultured at 20-day intervals. To induce shikonin production, Mpr cells were transferred to M9 liquid medium containing the same concentrations of sucrose and phytohormones, and cultured at 23 °C with shaking at 100 rpm in the dark for 20 days.

Axenic cultures of *L. erythrorhizon* shoots were grown on 1/2 Murashige and Skoog (MS) agar containing 1% (w/v) sucrose at pH 5.7 under a 16 h light/8 h dark photoperiod. Subcultures were performed at three-month intervals. Hairy root tissues of *L. erythrorhizon* were maintained on 1/2 MS agar medium containing 1% (w/v) sucrose and 20 mg l⁻¹ meropenem trihydrate and cultured at two-month intervals before being transferred to a 1/2 MS liquid medium containing the same concentrations of sucrose and meropenem trihydrate. They were then cultured at 100 rpm at 23 °C in the dark for 45 days. To induce shikonin production, hairy roots were cultured in M9 liquid medium containing 3% (w/v) sucrose without meropenem trihydrate in the dark.

Seeds of *Arabidopsis thaliana* (accession Columbia-0) were surface sterilized with 70% (v/v) ethanol and sown on MS medium (Wako) containing 1% (w/v) sucrose and 0.5% (w/v) gellan gum (Wako). The resulting *Arabidopsis* plants were grown on plates at 22 °C under continuous light for approximately one month. They were then transferred to peat moss (Jiffy-7, 32170142, Jiffy Products International AS, Stange, Norway) for further growth.

Generation of *L. erythrorhizon* hairy roots

Hairy roots of *L. erythrorhizon* were generated following the procedure described by Tatsumi and Ichino et al. (2020). Dedifferentiated leaves and stems excised from axenically cultured *L. erythrorhizon* shoots were then infected with *Rhizobium radiobacter* (Ri) strain A13 (MAFF 02-

10266) (Daimon et al. 1990; formerly referred to *Agrobacterium rhizogenes*) to induce hairy roots. The hairy roots generated by this procedure were grown and maintained on 1/2 MS agar medium (pH 5.7) containing 1% (w/v) sucrose and 20 mg l⁻¹ meropenem trihydrate. The inducible vector, pER8GW_LeABCG2-mGFP, was used to express *LeABCG2-mGFP*. It was introduced into *R. radiobacter* (strain A13) using the freeze-thaw transformation method. Transformed clones were obtained by selection with 100 mg l⁻¹ spectinomycin dihydrochloride pentahydrate and 50 mg l⁻¹ hygromycin B. Non-transformed A13 wild type and transformed A13 clones harboring the pER8GW_LeABCG2-mGFP vector were infected to *L. erythrorhizon* to generate hairy roots. These were then subjected to RT-qPCR analysis to assess the gene expression levels of *LeABCG1* and *LeABCG2*, and to perform membrane localization analysis for LeABCG2. Hairy roots infected with the A13 harboring vector pER8GW_LeABCG2-mGFP were selected on 1/2 MS agar medium containing 5 mg l⁻¹ hygromycin B, 1% (w/v) sucrose, and 20 mg l⁻¹ meropenem trihydrate to establish transformed lines to be used for further microscopic analysis.

RNA extraction and cDNA synthesis

For cDNA cloning, total RNA was extracted from *L. erythrorhizon* cells (strain T-TOM) cultured for 14 days in M9 liquid medium, and from 15-day-old *A. thaliana* seedlings. RNA extraction was performed using RNeasy Plant Mini Kit (74904, QIAGEN, Hilden, Germany). Extracted RNA samples were then treated with DNase using a DNA-free DNA Removal Kit (Ambion, Carlsbad, CA, USA) and RNA samples were then reverse transcribed into cDNAs using SuperScript III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA).

For RT-qPCR analysis, total RNA was extracted from *L. erythrorhizon* hairy roots grown in 1/2 MS and M9 liquid medium for 28 days in the dark using the RNeasy Plant Mini Kit. Total RNA samples from hairy roots treated with intracellular/extracellular metabolite extracts and standard samples were also prepared using the same method. RNA extracted from hairy roots, from cultured *L. erythrorhizon* cells, and from intact *L. erythrorhizon* tissue (Ueoka et al. 2020) was then treated to remove genomic DNA. Finally, cDNAs were prepared by reverse transcription using ReverTra Ace qPCR RT Master Mix with gDNA Remover (FSQ-301, Toyobo, Osaka, Japan).

Genomic DNA extraction

Genomic DNA was extracted from *L. erythrorhizon* cells (strain T-TOM) cultured on LS agar medium for 45 days as follows. First, cells were homogenized in a DNA extraction buffer (200 mM Tris-HCl [pH 8.0], 250 mM NaCl, 25 mM EDTA [pH 8.0], 0.5% [w/v] SDS) by grinding in a 1.5 mL tube (#96.07246.9.01, Treff Lab, Switzerland) using a micropestle (#96.07339.9.03, Treff Lab). After centrifugation of the homogenates at 20,400 g for 20 min, the supernatant was transferred to a new tube to which an equal volume of 2-propanol was added and mixed. The sample was centrifuged again

at 20,400 g for 10 min, and the pellet was washed with 70% (v/v) ethanol. The dried pellet was then redissolved in 1 × TE to obtain a genomic DNA solution.

Determination of cDNA and genomic sequences

To determine the full length of the cDNA sequence of *LeABCG1*, we first obtained the sequences upstream and downstream of comp46786_c2_seq1 for rapid amplification of both cDNA ends (i.e., 5'RACE and 3'RACE) using the cDNA derived from cultured *L. erythrorhizon* cells as a template for the SMARTer RACE 5'/3' Kit (634858, Clontech, Mountain View, CA, USA). All procedures were performed according to the manufacturer's protocol. The resulting sequences were then assembled with the contig sequence to predict the full-length cDNA sequence of *LeABCG1*. The full-length cDNA sequences of *LeABCG2* and *actin* (*LeACT*) genes for *L. erythrorhizon* were also predicted from the contig sequences comp78121_c0_seq10 and comp90574_c1_seq1, respectively.

The complete cDNA sequences of *LeABCG1* and *LeABCG2* containing in-frame stop codons before and after the ORF were amplified using KOD Plus Neo DNA Polymerase (KOD-401, Toyobo) using cDNA from cultured *L. erythrorhizon* cells as the template. The cDNA sequence of *LeACT* containing in-frame stop codons upstream to and downstream of the ORF was also amplified using PrimeSTAR Max DNA Polymerase (R045A, Takara, Kusatsu, Japan) from the cDNA of the cultured cells. The genomic sequences of these three genes, *LeABCG1*, *LeABCG2*, and *LeACT*, were then amplified using KOD Plus Neo and/or PrimeSTAR Max with genomic DNA from cultured *L. erythrorhizon* cells as the template. The fragments were amplified as above with adenine added to the 3' ends using A-attachment mix (TAK-301, Toyobo). These fragments were then subcloned into a pGEM-T Easy vector using T4 DNA ligase (A1360, Promega, Madison, WI, USA) to verify the insert sequences. All primers used for the amplification of cDNA and genomic fragments of *LeABCG1*, *LeABCG2*, and *LeACT* are listed in Table S1, and all constructed vectors are listed in Table S2.

Phylogenetic analysis

The amino acid sequences of *LeABCG1* and *LeABCG2* were deduced from the cDNA sequences determined as described above using the ExPASy Translate tool (<https://web.expasy.org/translate/>). The amino acid sequences of all ABCG proteins found in *A. thaliana* (i.e., from AtABCG1 to AtABCG43) and *Solanum lycopersicum* (i.e., from SlABCG1 to SlABCG70) were obtained from the latest versions of the public Arabidopsis and tomato databases, The Arabidopsis Information Resource (version TAIR10; <https://www.arabidopsis.org/>) and Sol Genomics Network (version SL4.00 and ITAG4.0; <https://solgenomics.net/>), respectively. All gene nomenclatures are based on previous studies (Garcia et al. 2004; Ofori et al. 2018; Verrier et al. 2008). The amino acid sequence of *Petunia x hybrida* PhABCG1 was obtained from UniProtKB (<https://www.uniprot.org/help/uniprotkb>). To obtain contigs encoding the half-size ABCG protein for *L. erythrorhizon*, we performed a homology

search using tBLASTn to query the half-size ABCG protein sequences from Arabidopsis (i.e., AtABCG1 to AtABCG28). For this search, we used *L. erythrorhizon* transcriptome data from a previous report (Takanashi et al. 2019) as a reference. We then selected 353 contigs without redundancy, with the cutoff set at e-20. The BLAST search was performed using BioEdit (version 7.2.5) (Hall 1999). The polypeptide sequences of each contig were then deduced using the ExpASy Translate tool. 57 contigs encoding half-sized ABCG proteins (including 28 non-redundant genes) from *L. erythrorhizon*, excluding LeABCG1 and LeABCG2, were selected from the 353 contig sequences according to the following criteria. The selected candidates (1) must belong to the half-size ABCG clade in the ABCG phylogenetic tree, (2) must contain a portion of the NBD domain starting at Walker A and ending at Walker B, and (3) must represent the highest expression level for transcript variants derived from the same gene.

Multiple sequence alignment was then performed by MUSCLE using the default settings for the NBD domain. Maximum likelihood trees were constructed using MEGA version 10.1.8 (Kumar et al. 2018; Stecher et al. 2020). The optimal phylogenetic model was selected using the model prediction tool implemented in MEGA and was used for all tree constructions. Initial trees for the heuristic search were automatically obtained by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with the superior log-likelihood value. A discrete gamma distribution was used to model evolutionary rate differences between sites (five categories, +G, parameter = 1.5837 for Fig. 1a and 1.7032 for Fig. S3). The tree was drawn to scale, with branch lengths measured by the number of substitutions per site. For all developed trees, classification as members of the ABCG subfamily and of class D was performed according to the methods of Zhang et al. (2021) and Panikashvili et al. (2011), respectively. All protein sequences used for phylogenetic analyses are listed in Table S3.

Prediction of NBD motifs and transmembrane helices

Characteristic motifs of the entire NBD domain were identified by multiple sequence alignment using BioEdit version 7.2.5 according to the protocols published by Linton (2007) and Cheepala et al. (2013). The number and position of transmembrane helices in the transmembrane domains (TMDs) of ABCG proteins were predicted using DeepTMHMM (<https://dtu.biolib.com/DeepTMHMM>) (Hallgren et al. 2022). These predictions were then verified and visualized using TMHMM version 2.0 (<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>) (Krogh et al. 2001; Sonnhammer et al. 1998). The percentage of identity between polypeptide sequences was calculated using full-length sequences of ABCG proteins or full NBD domains between the Walker A and D-loop motifs. All calculations were performed with Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

RT-qPCR

The qPCR primers amplifying the *LeABCG1*, *LeABCG2*, and *LeACT* transcripts were designed to anneal to the respective cDNA sequences. In addition, the amplified fragment of *LeABCG2* contained the 8th intron in the genomic sequence. The relative abundance of the *LeABCG1* and *LeABCG2* transcripts in the cDNA solution was then assessed by RT-qPCR, performed on a CFX96 Deep Well real-time system (Bio-Rad, Hercules, CA, USA) or LightCycler 96 (Roche, Basel, Switzerland) using Thunderbird SYBR qPCR Mix (QPS-201, Toyobo). The relative expression levels of *LeABCG1* and *LeABCG2* were calculated by the $\Delta\Delta C_t$ method using the *LeACT* gene used as an internal standard. All primers used for qPCR are listed in Table S1.

Metabolite extraction and treatment of hairy roots

L. erythrorhizon cells (strain Mpr) cultured in M9 medium were harvested and cell and medium samples were separated by filtration with Miracloth (475855, Merck, Darmstadt, Germany). Cell samples dried at 65 °C for two nights were homogenized in chloroform:methanol solution (2:1) using a mortar and a pestle to extract the lipophilic metabolites accumulated in the cells as “intracellular fractions”. Medium samples (30 mL) were partitioned with 15 mL of chloroform:methanol solution (2:1) followed by vigorous shaking to obtain the lipophilic metabolites secreted by the cells (extracellular fraction) from the organic phase. The solvent was evaporated under nitrogen stream at 60 °C to dry completely. The dried extracts were dissolved in dimethyl sulfoxide.

Hairy roots were cultured in fresh 1/2 MS liquid medium for 21 days. The hairy roots (500 mg fresh weight on average) were inoculated into 3.5 mL of 1/2 MS liquid medium supplemented with intracellular or extracellular crude metabolites at 3 mg/mL or 0.3 mg/mL, respectively, corresponding to the intracellular and extracellular concentrations *in vivo* (Fig. S6c, d). 1.2% (v/v) dimethyl sulfoxide was used as a mock in this experiment. The hairy roots were cultured at 23 °C at 50 rpm for 0 - 72 hours. For the treatment of standard samples, hairy roots (375 mg fresh weight on average) cultured in 1/2 MS liquid medium for 29 days were inoculated into 3.5 mL of 1/2 MS liquid medium supplemented with 1 mM acetylshikonin, 1 mM fatty alcohol (1-hexacosanol), 1 mM ω -hydroxy fatty acid (16-hydroxyhexadecanoic acid), 1 mM monoacylglycerol (2-monopalmitin), and 0.1 mM triacylglycerol (1,2-palmitin-3-palmitolein) at each final concentration. Each applied concentration corresponds to 0.2-fold (acetylshikonin), five-fold (1-hexacosanol), 0.1-fold (2-monopalmitin), and 100-fold (1,2-palmitin-3-palmitolein) compared to that of shikonin derivatives secreted from Mpr cells (Nakanishi et al. 2024), that of fatty alcohols as suberin monomers reported in Arabidopsis (Shanmugarajah et al. 2019), that of monoacylglycerols as cutin precursors in tomato fruits (Elejalde-Palmetto et al. 2021), and that of triacylglycerols reported to be secreted from T-TOM cells (Tatsumi et al. 2023). 2.0% (v/v) dimethyl sulfoxide was used as a control. The hairy roots were cultured for 72 hours. Statistical analyses were performed using Tukey HSD test ($\alpha = 0.05$) after one-way ANOVA analysis.

Vector construction

Complete ORF sequences with and without a stop codon of *LeABCG1* cDNA taken from the cDNA pool of cultured *L. erythrorhizon* cells were amplified using KOD Plus Neo. The amplified fragments were then individually inserted into pENTR/D-TOPO vectors by the TOPO reaction (K240020, Thermo Fisher Scientific, Waltham, MA, USA) to obtain entry clones. The ORF of the *LeABCG2* gene with and without a stop codon in the T-TOM cDNA was also amplified using KOD Plus Neo and then cloned into a linearized pENTR/D-TOPO vector using the In-Fusion HD Cloning Kit (639648, Clontech). The *LeABCG1* and *LeABCG2* ORF fragments in the entry clones were then transferred into the destination vectors pGWB406m and pGWB405m (Nakagawa et al. 2007; Segami et al. 2014), using LR clonase II enzyme mix (11791100, Thermo Fisher Scientific). This resulted in the following plant expression vectors: pGWB406m_LeABCG1 (mGFP-LeABCG1), pGWB405m_LeABCG1 (LeABCG1-mGFP), pGWB406m_LeABCG2 (mGFP-LeABCG2), and pGWB405m_LeABCG2 (LeABCG2-mGFP).

LeABCG2-mGFP in pGWB405m_LeABCG2 was amplified using KOD Plus Neo and cloned into a linearized pENTR/D-TOPO vector by the In-Fusion reaction. The Gateway cassette of pGWB502 (Nakagawa et al. 2007) was also amplified using KOD Plus Neo before being inserted into a pER8 vector (Zuo et al. 2000), digested with *Xho* I and *Spe* I, and then subjected to the In-Fusion reaction to construct a pER8GW vector. The *LeABCG2-mGFP* fragment was then inserted into the destination vector pER8GW by LR recombination, resulting in the plasmid pER8GW_LeABCG2-mGFP.

The coding sequence, containing a stop codon, of the Arabidopsis *AtPIP1;4* gene was amplified with KOD Plus Neo using a WAVE 138Y plasmid (Geldner et al. 2009) as a template. The resulting amplicon was then inserted into a linearized pENTR/D-TOPO via In-Fusion reaction. Next, the cDNA coding sequence, containing a stop codon, of the Arabidopsis *VAM3* gene was amplified using PrimeSTAR Max and inserted into a pENTR/D-TOPO vector via TOPO cloning. The resulting insertions were then transferred into the destination vector pGWB555 (Nakagawa et al. 2007) by LR recombination to yield the expression vectors, pGWB555_AtPIP1;4 (mRFP-AtPIP1;4) and pGWB555_AtVAM3 (mRFP-AtVAM3). All primers used for vector construction and all constructed vectors are listed in Tables S1 and S2, respectively.

Transient expression of foreign genes in *Nicotiana benthamiana*

Vector constructs for plant gene expression, i.e., LeABCG1-mGFP, mGFP-LeABCG1, LeABCG2-mGFP, mGFP-LeABCG2, mRFP-AtPIP1;4, and mRFP-AtVAM3, were then introduced into *R. radiobacter* (Ti) strain LBA4404 using the freeze-thaw method. After confirmation of the desired insertion into agrobacteria clones by PCR, the transformed clones were cultured in YEB liquid medium supplemented with 50 mg l⁻¹ streptomycin sulfate and 100 mg l⁻¹ spectinomycin

1
2
3 360 dihydrochloride pentahydrate. The overnight cultures were centrifuged at 20,400 g for one minute,
4 361 and then the pellets were washed thrice and resuspended in distilled water. Infiltration solutions were
5 362 prepared at $OD_{600} = 0.2$ for expression of the GFP-fused ABCG protein, or at $OD_{600} = 0.1$ for
6 363 colocalization analysis. The agrobacteria solutions were infiltrated into the expanded leaves of five-
7 364 week-old *N. benthamiana* plants to express GFP fusion proteins.
8
9 365

10 366 **Confocal microscopic analysis**

11
12 367 The hairy roots harboring the *LeABCG2-mGFP* gene were treated with 20 μ M β -estradiol or dimethyl
13 368 sulfoxide (final concentration: 0.2% [v/v]) in 1/2 MS liquid medium for 20 hours to express the
14 369 LeABCG2-mGFP fusion protein. *N. benthamiana* leaves at two days after infection and the *L.*
15 370 *erythrorhizon* hairy roots were observed using a FV3000 confocal laser scanning microscope
16 371 (Olympus, Tokyo, Japan) with a 60×0.75 numerical aperture water immersion objective (UPLSAPO
17 372 60XW, Olympus). GFP fluorescence was monitored by excitation at 488 nm with a 20-mW diode laser
18 373 and emission at 500–540 nm; RFP fluorescence was detected by excitation at 561 nm with a 20-mW
19 374 diode laser and emission at 570–620 nm; plastid autofluorescence was detected by excitation at 640
20 375 nm with a 40-mW diode laser and emission at 650–750 nm. Image acquisition and analysis were
21 376 performed using FV31S-SW software (Olympus).
22
23 377

24 378 **Results**

25 379 **Screening of membrane transporter genes in *Lithospermum erythrorhizon***

26 380 To identify membrane transporters that preferentially export lipidic compounds in *L. erythrorhizon*,
27 381 which produces divergent lipophilic metabolites such as shikonin derivatives, shikono/echinofurans,
28 382 and triacylglycerol, we first screened RNA-seq data for *L. erythrorhizon* (Takanashi et al. 2019). This
29 383 transcriptome data consisted of three types of paired plant samples; root tissues from intact *L.*
30 384 *erythrorhizon* plants (i.e., shikonin-producing root bark and the peeled root that does not produce
31 385 shikonin), hairy roots grown in M9 liquid medium either under illumination or in the dark, and cells
32 386 cultured in M9 liquid medium also under light or in the dark. This is because light strongly suppresses
33 387 shikonin production even in the shikonin production medium, M9. According to the gene expression
34 388 levels calculated in fragments per kb per million mapped reads (FPKM), we identified 487 contigs
35 389 that were at least 3-fold upregulated in all three pairs under shikonin-producing conditions. Of these,
36 390 23 contigs, including 20 non-redundant genes, were annotated as membrane transporters. These
37 391 included several transporter families (Table S4), out of which we focused on one contig
38 392 comp46786_c2_seq1, which was predicted to code for a G subfamily of ABC transporters (ABCG),
39 393 as members of this subfamily are often associated with the secretion of lipophilic metabolites in plants
40 394 (Ichino and Yazaki 2022). This contig appeared to be expressed exclusively in shikonin-producing
41 395 conditions based on comparisons of all three tissue pairs (Table S4). Because this comp46786_c2_seq1

had only a partial open reading frame (ORF) (Fig. S1), we performed rapid amplification of cDNA ends (5'RACE and 3'RACE), to obtain a cDNA sequence containing the complete ORF and an in-frame stop codon at the 5'-untranslated region. We designated this gene as *LeABCG1*.

We selected an additional ABCG protein, because the expression level of *LeABCG1* was moderate and the highest FPKM value among the six samples was only about 10 in the root bark (Table S4). A tBLASTn search using the PhABCG1, which is a half-size ABCG transporter from *Petunia x hybrida* involved in the export of lipophilic methylbenzoate and benzyl alcohol (Adebesin et al. 2017), as a query highlighted six genes of *L. erythrorhizon* showing similarity to PhABCG1 (at a cutoff e-value of e-70; see Table S5). Among them, the one revealing the highest gene expression, comp78121_c0_seq10, was selected as the second candidate (Fig. S1; Table S6). After cloning of the cDNA sequence, it was designated as *LeABCG2*.

LeABCG1 and LeABCG2 belong to Class A and D of half-size ABCG proteins, respectively

The ORF sequences of the *LeABCG1* and *LeABCG2* genes were predicted to be 2,253 bp and 2,172 bp, respectively, based on cDNA sequences. Their putative polypeptide sequences were 750 and 723 amino acids, respectively, and both were predicted to be half-size ABCG proteins, previously termed the white-brown complex (Fig. S2). To gain insight into the phylogenetic position of the *LeABCG1* and *LeABCG2* proteins within the plant ABCG subfamily, we then performed a phylogenetic analysis. First, we performed a homology search with tBLASTn using all 28 Arabidopsis half-size ABCG protein sequences as queries and found 30 genes encoding half-size ABCG proteins in the *L. erythrorhizon* transcriptome data, including *LeABCG1* and *LeABCG2* (Table S6). The resulting phylogenetic tree consisted of seven groups, all of which were monophyletic (Figs. 1a, S3), and classified from A to G according to Zhang et al. in Arabidopsis (Zhang et al. 2021). Half-size ABCG proteins are grouped in clades A to F, whereas all full-size ABCG proteins, also referred to as pleiotropic drug resistance (PDR) proteins, belong exclusively to class G (Figs. 1a, S3). In this phylogenetic tree, *LeABCG1* belongs to class A, whereas *LeABCG2* belongs to class D (Fig. 1a).

The class A subfamily includes eight Arabidopsis ABCGs, six tomato ABCGs, and at least five *L. erythrorhizon* ABCGs (Fig. 1a). Two *L. erythrorhizon* genes, *LeABCG1* and comp55537_c0, are orthologous to *AtABCG2* and *AtABCG20*, which have been shown to function in suberin deposition and pollen exine formation (Fedi et al. 2017; Yadav et al. 2014). Our phylogenetic analysis suggests that *LeABCG1* may recognize suberin monomers as substrates to export them to the extracellular space. In addition, some Arabidopsis members such as *AtABCG6*, *AtABCG16*, and *AtABCG20* are involved in the translocation of jasmonic acid (Li et al. 2017; Li et al. 2020). Furthermore, the involvement of *AtABCG1* and *AtABCG16* in auxin signaling and of *AtABCG17* and *AtABCG18* in abscisic acid translocation have been reported (Liu et al. 2020; Zhang et al. 2021). The possible involvement of *LeABCG1* in the export of suberin monomers and pollen wall components, as well as

in the transport of other metabolites (e.g., phytohormones) will be discussed later.

The class D ABCG subfamily, to which *LeABCG2* belongs, consists of five Arabidopsis ABCGs, nine tomato ABCGs, and at least four *L. erythrorhizon* ABCGs (Fig. 1a). Members of this subfamily are further divided into several subgroups, clades I-IV (Panikashvili et al. 2011), and some Arabidopsis members are involved in cuticle formation (Elejalde-Palmett et al. 2021; McFarlane et al. 2010; Pighin et al. 2004). *LeABCG2* was grouped within clade I (Fig. 1a). Given the conservation of function of class D members in cuticle deposition, the possible involvement of *LeABCG2* in the secretion of hydrophobic materials in *L. erythrorhizon* tissues will be discussed later.

The sequence identities of the NBD region between *LeABCG1* and its orthologs in Arabidopsis and tomato were both about 90%, whereas those of the entire *LeABCG1* sequence to other orthologs were about 70% (Table S7a, b). The polypeptide sequence identities between *LeABCG2* and other orthologs were also very high (Table S7c, d). The NBD region contains several conserved motifs, and *LeABCG1* possesses these conserved motifs, i.e., Walker A (GASGSGKST), Q-loop (MQD), ABC signature (VSGGE), Walker B (II/LLFLDE), and D-loop (LD) (Fig. 1b). In *LeABCG2*, Q-loop and H-loop adding to an aromatic motif (Y) were not detected, whereas other motifs, including the Walker A (GPAKSGKST), ABC signature (LS/RSGE), Walker B (HA/ILFIDE), and D-loop (LD), are conserved (Fig. 1c).

***LeABCG1* is an intron-less gene while *LeABCG2* contains multiple introns**

To further analyze *LeABCG1* and *LeABCG2* in detail, we compared the genomic structures and positions of each motif of the NBD and transmembrane α -helices of *LeABCG1* and *LeABCG2* with those of their orthologs in other plant species. We first obtained genomic DNA for *LeABCG1* and *LeABCG2* from cultured *L. erythrorhizon* cells and determined their genomic structures. To our surprise, we found that *LeABCG1* (2,364 bp in length) contained no intron in its ORF region (Fig. 2a), while the *LeABCG2* gene (6,469 bp in length) was composed of 11 exons and 10 introns (Fig. 2b). Further analysis revealed that all *LeABCG1* orthologs in other plants were intronless, except for *SLABCG13*, which contains a single intron located upstream of the ABC signature motif (Fig. 2a, c). In addition, these genes commonly contain six transmembrane α -helices with conserved genomic locations. The first five α -helices are organized in a compact bundle and the sixth helix, near the C-terminus, is connected by a long loop (Figs. 2a, S2). The structure *L. erythrorhizon LeABCG1* is highly conserved with its Arabidopsis and tomato orthologs. Furthermore, all other *LeABCG1* orthologs of the class A ABCGs in the Arabidopsis genome possess intronless ORFs (Fig. S4). In tomato, *SLABCG8*, *SLABCG9*, and *SLABCG23* were also intronless. Consistently, we found that most class A ABCG members in *Brassica rapa*, pineapple (*Ananas comosus*), grapevine (*Vitis vinifera*), and white lupin (*Lupinus albus*), were either intronless or possessed only a single intron (Aslam et al. 2021; Chen et al. 2017; Tang et al. 2018; Yan et al. 2017). These results suggest that the common ancestor of these

class A ABCG proteins had no introns, and that this feature was conserved during plant evolution. Regarding the *LeABCG2* orthologs, the ORF regions of *AtABCG3* and *SlABCG28* also contained 11 exons and 10 introns (Fig. 2b). In these genes, the exon lengths of the 2nd to 10th exons were identical to those found in *LeABCG2*, while the intron lengths were comparatively variable (Fig. 2b). The exon positions of each motif in the NBD domain and of the six transmembrane α -helices were identical among all *LeABCG2* orthologs (Fig. 2b, d). Specific positions of exon-intron junctions were also conserved: for example, the 1st intron of the ORF region was inserted 19 bp downstream of the Walker A motif (Fig. S5a), the 3rd intron of the ORF region was inserted 2 bp after the D-loop motif (Fig. S5b), and the 9th intron of the ORF region interrupted the coding sequence of the fifth transmembrane helix (Fig. S5c). Taken together, we conclude that the genomic structures of ABCG proteins are highly conserved among orthologous genes in dicotyledonous plants.

Tissue- and cell-specific expression of *LeABCG1* and *LeABCG2*

As described above, we suggested that *LeABCG1* expression was inducible under shikonin-producing conditions based on the FPKM values of the transcriptome analysis (Fig. 3a). In contrast, *LeABCG2* appeared to be constitutively and highly expressed in all six tissue samples (Fig. 3b). To investigate the actual expression patterns of *LeABCG* genes in *L. erythrorhizon* samples, we performed reverse transcription quantitative PCR (RT-qPCR) using gene-specific primers. In both cultured cells and hairy root cultures, *LeABCG1* expression was not induced under shikonin-producing conditions, i.e., in M9 medium in the dark (Fig. 3c, d); this was inconsistent with transcriptome data (Fig. 3a). In all *L. erythrorhizon* plants, we observed high levels of *LeABCG1* expression in above-ground tissues, including flowers, leaves, and stems, i.e., where shikonin biosynthesis does not occur under normal growth conditions under illumination (Fig. 3e). However, in root tissues, we observed a 13-fold higher expression of *LeABCG1* in shikonin-producing root bark than in peeled non-shikonin-producing roots. Even so, the level of *LeABCG1* expression in root bark was 30-fold lower than in flowers and leaves (Fig. 3e). Regarding the potential roles of *LeABCG1* as a class A member (Fig. 1a), *LeABCG1* may play a role in the export of suberin monomers and pollen wall components, and may also play a role in the transport of other metabolites (e.g., phytohormones) from epidermal cells to tissue surfaces in both roots and aerial parts of the plant.

LeABCG2 was found to be slightly upregulated in both cultured cells and hairy roots grown in M9 medium (Fig. 3f, g). Furthermore, while previous studies have shown that light exposure shuts down shikonin production (Yazaki 2017), we found little light effect on *LeABCG2* expression in cultured cells (Fig. 3f). In intact plants, *LeABCG2* was constitutively expressed in all tissues examined in this study, with slightly higher expression levels in stems and in roots compared to other tissues (Fig. 3h). These data suggest that *LeABCG2* plays a common role independent of differences associated with culture conditions or plant tissues. Considering the conservation of function of class D members in

cuticle deposition as mentioned above (Fig. 1a), LeABCG2 may also be involved in the secretion of hydrophobic materials from plant cells.

Response of *LeABCG1* and *LeABCG2* expression to the endogenous *L. erythrorhizon* metabolites

The expression of some ABCG transporter genes is often regulated by the presence of transport substrates. For example, *ABCG20* in *V. vinifera* is up-regulated in the response to suberin monomers (Wang et al. 2023). To gain insight into the potential transport substrates of two LeABCG transporters, hairy roots were treated with the endogenous metabolites of *L. erythrorhizon*. The influence on the gene expression of *LeABCG1* and *LeABCG2* was evaluated by RT-qPCR. The endogenous metabolites were separated into two samples, i.e., the extract prepared from cultured *L. erythrorhizon* cells and the extracellular metabolites recovered from the culture medium, which were then added to the hairy root cultures (Fig. S6a-d). In the mock control, *LeABCG1* expression was gradually decreased (Fig. 4a) to a low expression level of *LeABCG1* at steady state as shown in Fig. 3a. This profile suggests that the *LeABCG1* expression is sensitive to environmental stress such as passaging culture shock from old to fresh culture media. When treated with either intracellular or extracellular extract, *LeABCG1* expression was also decreased over the culture period, with a tendency to decrease *LeABCG1* expression compared to the control. Specifically, a slight reduction was observed in the intracellular extract at four to ten hours after treatment, while a slightly greater reduction was observed in the extracellular extract at 24 hours and 72 hours than in the intracellular extract (Fig. 4a). In contrast, *LeABCG2* expression in the hairy roots was gradually increased during 72-hours of culture in both sham and extract treatment groups (Fig. 4b). Significantly increased expression of *LeABCG2* was found by extracellular extract treatment at 24 and 72 hours (Fig. 4b). In addition, intracellular extract treatment also slightly increased *LeABCG2* expression at the same time points (Fig. 4b).

To investigate which type of metabolites affect the gene expression of *LeABCGs* as putative transport substrates, candidate compounds were tested individually. We selected five compounds from shikonin derivatives and lipophilic metabolites that accumulate in the apoplast, namely, acetylshikonin (representative shikonin derivative), a fatty alcohol (suberin monomer), ω -hydroxy fatty acid and monoacylglycerol (cutin monomers), and triacylglycerol (Fig. S6e). However, these five apoplastic metabolites and monomers had little effect on the expression of *LeABCG1* or *LeABCG2* (Fig. 4c, d), whereas treatment with acetylshikonin and ω -hydroxy fatty acid showed a tendency to increase *LeABCG1* expression without reaching statistical significance (Fig. 4c). In addition, *LeABCG2* expression was slightly increased by acetylshikonin and slightly decreased by ω -hydroxy fatty acid (Fig. 4d). Contrary to the reports showing a positive correlation between gene expression and the addition of its transport substrate, there are ABC transporter genes that are repressed when excess amounts of transport substrates are present. *LeABCG2* belongs to the former group and *LeABCG1* might be a member of the latter. Unfortunately, the transport substrates of these LeABCGs could not

be identified in this study, although we found that the presence of endogenous metabolites of *L. erythrorhizon* may regulate the transport function of LeABCG1 and LeABCG2.

Membrane localization of LeABCG1 and LeABCG2

To study the subcellular localization of LeABCG1 and LeABCG2 proteins, we expressed monomeric GFP-fused LeABCG proteins. We developed fusion proteins tagged with GFP at both the C-terminal and N-terminal ends, i.e., both LeABCG-mGFP and mGFP-LeABCG. Transient expression of these GFP fusion proteins in *N. benthamiana* leaves showed that both LeABCG1-mGFP and mGFP-LeABCG1 were distributed along the outlines of leaf epidermal cells (Figs. 5a, S7a, b). This indicated that LeABCG1 was localized to the PM. GFP fluorescence signals for LeABCG2-mGFP and mGFP-LeABCG2 were also detected along the outlines of leaf epidermal cells (Figs. 5a, S7a, b). The distribution of the LeABCG2 protein on the cell membrane was also confirmed by examination of *L. erythrorhizon* hairy roots expressing LeABCG2-mGFP (Fig. S8).

To determine the identity of the membrane on which these two LeABCG proteins are localized, we co-expressed LeABCG-mGFP with mRFP-fused Arabidopsis plasma membrane intrinsic protein 1;4 (mRFP-AtPIP1;4) or mRFP-fused Arabidopsis VACUOLAR MORPHOLOGY 3 (mRFP-AtVAM3), which are markers of the PM and vacuolar membrane, respectively (Geldner et al. 2009; Uemura et al. 2002). LeABCG1-mGFP colocalized completely with mRFP-AtPIP1;4 but not with mRFP-AtVAM3 (Figs. 5b, S7c). The GFP signal was visible between two vacuolar membranes labeled with mRFP-AtVAM3 in two adjacent cells (Fig. 5b). A similar pattern was observed in LeABCG2-mGFP-expressing cells (Figs. 5c, S7d). Based on these findings, we concluded that both LeABCG1 and LeABCG2 proteins localize to the PM.

Interestingly, the GFP signals of LeABCG2-mGFP and mGFP-LeABCG2 were observed as punctate structures as well as linear patterns along the PM, whereas LeABCG1 was uniformly distributed along the PM (Fig. 5a, bottom panels). Further examination revealed that the punctate structures associated with LeABCG2 were located on the PM (Fig. 5c). There are several possible explanations for the formation of punctate structures associated with LeABCG2. One is that the LeABCG2 protein may form supramolecular complexes as well as dimers, thereby resulting in LeABCG2 clusters on the PM. In addition, the micro-distribution of LeABCG2 on the PM may be susceptible to membrane microdomains, such as lipid rafts, that concentrate specific membrane proteins and membrane lipids to form unique and heterogeneous lipid compositions that contribute to the functional compartmentalization of the membrane. For example, several reports have shown that the localization and function of human ABC transporters are dependent on lipid composition and membrane rafts (Bacso et al. 2004; Kok et al. 2014). Therefore, LeABCG2 may be distributed in specific microdomains on the PM, or LeABCG2 itself may induce the formation of LeABCG2-specific microdomains.

Discussion

To our knowledge, LeABCG1, and LeABCG2 are the first two ABCG transporters isolated from shikonin-producing boraginaceous plants. Although the gene expression of *LeABCG1* and *LeABCG2* varies with tissue and culture conditions, *L. erythrorhizon* plants and cells always secrete lipophilic metabolites, including wax, cutin, suberin and triacylglycerol, as well as shikonin derivatives. In *L. erythrorhizon*, in addition to the ABCG transporters, two membrane transporter genes have been previously isolated: *LeMDR* and *LeMRP* (Zhu et al. 2017, 2018). The former is classified as a B-type ABC transporter (ABCB), while the latter is classified as a C-type ABC transporter (ABCC). Both ABC transporter genes are reported to be expressed in shikonin-producing samples (i.e., root tissues and in M9 cultured cells of *L. erythrorhizon*). Furthermore, the downregulation of either gene by RNAi resulted in reduced shikonin production in hairy roots of *L. erythrorhizon* (Zhu et al. 2017, 2018). Plant ABCC transporters localize either along the PM or in tonoplasts, and mediate the translocation of transport substrates across membranes, showing a preference for hydrophilic substances such as anthocyanins and crocins (Demurtas et al. 2019; Francisco et al. 2013). The LeMRP protein may be involved in the transport of unidentified hydrophilic metabolites into vacuoles, and its relationship to shikonin, which is secreted across the PM, is unknown (Zhu et al. 2018).

The *L. erythrorhizon* genome has recently been sequenced and analyzed using several different plant strains (Auber et al. 2020; Tang et al. 2021). *In silico* analysis using artificially combined genomes derived from two different strains of *L. erythrorhizon*, i.e., in which seeds originate from German institutions and Chinese companies, respectively (Auber et al. 2020; Tang et al. 2021), predicted that the *L. erythrorhizon* genome contains 118 ABC protein genes and 31 half-size ABCG members (Li et al. 2022). These data are largely consistent with our results, which revealed at least 30 half-size ABCG genes in a Japanese strain (Fig. 1a; Table S6). In addition, we found three additional half-size class A ABCG genes, including *LeABCG1* and its ortholog comp55537_c0, both of which are expressed in the Japanese strain but were not identified in the most recent version of the genome (Li et al. 2022).

Recent transcriptome analyses of boraginaceous plants producing shikonin, and alkannin derivatives, have listed candidate enzymes that may be related to shikonin biosynthesis (Ahmad et al. 2022; Rai et al. 2018; Suttiyut et al. 2022; Takanashi et al. 2019; Tang et al. 2020; Wu et al. 2017). Recently, intensive biochemical research on enzymes related to the shikonin biosynthetic pathway has identified several enzymes in *L. erythrorhizon*, including cytosolic GPP synthase (GPPS), cytochrome P450 monooxygenases, CYP76B100, CYP76B101, and CYP82AR2, and the acetyltransferase LeSAT1 (Oshikiri et al. 2020, 2024; Song et al. 2020, 2021; Ueoka et al. 2020) as well as CYP76B74 in *Arnebia euchroma* (Wang et al. 2019). Downregulation of these candidate genes in hairy roots also helped to identify the genes causally involved in shikonin biosynthesis *in planta*, including *LeGPPS*, the membrane-bound prenyltransferase *LePGT1*, *AeCYP76B74* and its ortholog in *Echium plantagineum*,

and *LeSAT1* (Auber et al. 2020; Fu et al. 2020; Suttiyut et al. 2022; Wang et al. 2019; Wang et al. 2022).

Understanding of the transport machinery responsible for the secretion of lipophilic metabolites into the extracellular space is another important research goal, especially since the secretion machinery remains less understood than the biosynthetic enzymes. This is partly due to technical difficulties associated with the biochemical characterization of G-type ABC transporters (Gräfe and Schmitt 2021). Nevertheless, with recent improvements in the methods used for heterologous expression of plant ABCG transporters (Gräfe et al. 2019), several studies have succeeded in demonstrating the transport activity of lipophilic metabolites (Adebesin et al. 2017; Elejalde-Palmett et al. 2021). Therefore, these transport assays may help us to determine the physiological transport substrates of LeABCG1 and LeABCG2 in future studies. In addition, loss-of-function technologies, such as gene silencing by hairpin RNA expression, virus-induced gene silencing, or genome editing of shikonin-producing plants, are also currently under development (Auber et al. 2020; Izuishi et al. 2020; Nakanishi et al. 2024; Wang et al. 2022; Zhu et al. 2017). Such technologies would allow us to understand when and in which tissues or cell types LeABCG1 and LeABCG2 function and can help characterize the physiological roles that both LeABCG proteins play in *L. erythrorhizon*. Therefore, technological advances will provide insight into the transport machinery responsible for the secretion of lipophilic metabolites from plant cells in the near future.

Supplementary Information

The online version contains supplementary material.

Acknowledgments

We are grateful to Dr. Hirobumi Yamamoto (Toyo University, Japan), Dr. Kazuyoshi Terasaka (Nagoya City University), and Dr. Noboru Onishi (Kirin Holdings Company Limited, Japan) for sharing cultured cells strains T-TOM, and Mpr, and axenic shoots of *L. erythrorhizon*, respectively. We also thank Dr. Masahiro Mii and Dr. Tomoko Igawa (Chiba University, Japan) for sharing *R. radiobacter* strain A13. We are grateful to Dr. Shoji Segami (National Institute for Basic Biology, Japan), Dr. Masayoshi Maeshima (Chubu University, Japan), and Dr. Tsuyoshi Nakagawa (Shimane University, Japan) for providing vectors pGWB405m and pGWB406m; to Dr. Tsuyoshi Nakagawa for providing vectors pGWB502 and pGWB555; and to Dr. Takashi Aoyama (Kyoto University, Japan) for providing vector pER8. We also thank the Arabidopsis Biological Resource Center and the Nottingham Arabidopsis Stock Centre for providing *A. thaliana* seeds and the WAVE 138Y plasmid, respectively. We are grateful to Dr. Akifumi Sugiyama and Dr. Ryosuke Munakata (Kyoto University) for fruitful discussion and preparation of chemical treatment, and to Dr. Nobukazu Shitan and Dr. Yasuyuki Yamada (Kobe Pharmaceutical University, Japan) for the chemical treatment experiment facility.

Finally, we thank the Development and Assessment of Sustainable Humanosphere system, RISH, Kyoto University, for supporting the plant culture facility.

Author contributions

T.I., K.Tat., and K.Y. conceptualized and designed the research. T.I., K.Tat., Y.M., and K.Tak. selected candidate genes. T.I., K.Tat., Y.M., and E.M. performed cloning and vector construction. T.I., K.Tat., and A.T. performed microscopic analysis and RT-qPCR. T.I., K.Tat., and M.N. performed phylogenetic analysis. T.I., K.Tat., and K.S. generated hairy root cultures. T.I. and K.Tat. prepared figures and tables. T.I., K.Tat., and K.Y. wrote the original manuscript. All authors approved the final version of the manuscript.

Funding

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI grants [Grant numbers JP19H05638 to K.Y., JP21K14828 and JP24K08837 to T.I.), a JSPS Research Fellowship [Grant number JP18J11502 to K.Tat.], and by a New Energy and Industrial Technology Development Organization (NEDO) grant [Grant number 16100890-0 to K.Y.]. This work was also supported by RIKEN Cluster for Science, Technology and Innovation Hub (RCSTI to K.Y.). Additional support was provided by the Institute for the Future of Human Society of Kyoto University for 2023 (to K.Y.), and by Humanosphere Science Research to T.I. and Mission Research 5 of the Research Institute for Sustainable Humanosphere (RISH) at Kyoto University.

Data availability

Accession numbers for raw RNA-seq data used in this study are DRX100780-DRX100785 deposited in Sequence Read Archive. Accession numbers for the genes mentioned in this study, except for those used in phylogenetic analysis, are present in the DDBJ/GenBank/EMBL or Arabidopsis Genome Initiative databases as follows: *LeABCG1* cDNA, LC781545; *LeABCG1* genomic DNA, LC781546; *LeABCG2* cDNA, LC781613; *LeABCG2* genomic DNA, LC781617; *LeACT* cDNA, LC781674; *LeACT* genomic DNA, LC781675; PhABCG1 protein, H9BZ66.1; AtPIP1;4, At4g00430; AtVAM3, At5g46860. Other data underlying this article will be shared upon reasonable request to the corresponding author.

Declarations

Competing interests

The authors have no conflicts of interest to declare.

References

- Adebesin F, Widhalm JR, Boachon B, Lefèvre F, Pierman B, Lynch JH, Alam I, Junqueira B, Benke R, Ray S, Porter JA, Yanagisawa M, Wetzstein HY, Morgan JA, Boutry M, Schuurink RC, Dudareva N (2017) Emission of volatile organic compounds from petunia flowers is facilitated by an ABC transporter. *Science* 356:1386–1388. <https://doi.org/10.1126/science.aan0826>
- Ahmad M, Varela Alonso A, Koletti AE, Rodić N, Reichelt M, Rödel P, Assimopoulou AN, Paun O, Declerck S, Schneider C, Molin EM (2022) Dynamics of alkannin/shikonin biosynthesis in response to jasmonate and salicylic acid in *Lithospermum officinale*. *Sci Rep* 12:17093. <https://doi.org/10.1038/s41598-022-21322-0>
- Aslam MM, Waseem M, Zhang Q, Ke W, Zhang J, Xu W (2021) Identification of ABC transporter G subfamily in white lupin and functional characterization of *L.albABGC29* in phosphorus use. *BMC Genomics* 22:723. <https://doi.org/10.1186/s12864-021-08015-0>
- Assimopoulou AN, Boskou D, Papageorgiou VP (2004) Antioxidant activities of alkannin, shikonin and *Alkanna tinctoria* root extracts in oil substrates. *Food Chem* 87:433–438. <https://doi.org/10.1016/j.foodchem.2003.12.017>
- Auber RP, Suttiyut T, McCoy RM, Ghaste M, Crook JW, Pendleton AL, Widhalm JR, Wisecaver JH (2020) Hybrid de novo genome assembly of red gromwell (*Lithospermum erythrorhizon*) reveals evolutionary insight into shikonin biosynthesis. *Horticul Res* 7:82. <https://doi.org/10.1038/s41438-020-0301-9>
- Bacso Z, Nagy H, Goda K, Bene L, Fenyvesi F, Matkó J, Szabó G (2004) Raft and cytoskeleton associations of an ABC transporter: P-glycoprotein. *Cytometry Part A* 61:105–116. <https://doi.org/10.1002/cyto.a.20081>
- Bird D, Beisson F, Brigham A, Shin J, Greer S, Jetter R, Kunst L, Wu X, Yephremov A, Samuels L (2007) Characterization of Arabidopsis ABCG11/WBC11, an ATP binding cassette (ABC) transporter that is required for cuticular lipid secretion. *Plant J* 52:485–498. <https://doi.org/10.1111/j.1365-313X.2007.03252.x>
- Brigham LA, Michaels PJ, Flores HE (1999) Cell-specific production and antimicrobial activity of naphthoquinones in roots of *Lithospermum erythrorhizon*. *Plant Physiol* 119:417–428. <https://doi.org/10.1104/pp.119.2.417>
- Cheepala S, Hulot JS, Morgan JA, Sassi Y, Zhang W, Naren AP, Schuetz JD (2013) Cyclic nucleotide compartmentalization: contributions of phosphodiesterases and ATP-binding cassette transporters. *Annu Rev Pharmacol Toxicol* 53:231–253. <https://doi.org/10.1146/annurev-pharmtox-010611-134609>
- Chen P, Li Y, Zhao L, Hou Z, Yan M, Hu B, Liu Y, Azam SM, Zhang Z, Rahman Z, Liu L, Qin Y (2017) Genome-Wide Identification and Expression Profiling of ATP-Binding Cassette (ABC) Transporter Gene Family in Pineapple (*Ananas comosus* (L.) Merr.) Reveal the Role of *AcABCG38* in Pollen Development. *Front Plant Sci* 8:2150.

- <https://doi.org/10.3389/fpls.2017.02150>
- Daimon H, Fukami M, Mii M (1990) Hairy root formation in peanut by the wild type strains of *Agrobacterium rhizogenes*. Plant Tissue Culture Letters 7:31–34. <https://doi.org/10.5511/plantbiotechnology1984.7.31>
- Demurtas OC, de Brito Francisco R, Diretto G, Ferrante P, Frusciante S, Pietrella M, Aprea G, Borghi L, Feeney M, Frigerio L, Coricello A, Costa G, Alcaro S, Martinoia E, Giuliano G (2019) ABCC Transporters Mediate the Vacuolar Accumulation of Crocins in Saffron Stigmas. Plant Cell 31:2789–2804. <https://doi.org/10.1105/tpc.19.00193>
- Elejalde-Palmett C, Martinez San Segundo I, Garroum I, Charrier L, De Bellis D, Mucciolo A, Guerault A, Liu J, Zeisler-Diehl V, Aharoni A, Schreiber L, Bakan B, Clausen MH, Geisler M, Nawrath C (2021) ABCG transporters export cutin precursors for the formation of the plant cuticle. Curr Biol 31:2111–2123. <https://doi.org/10.1016/j.cub.2021.02.056>
- Fedi F, O'Neill CM, Menard G, Trick M, Dechirico S, Corbineau F, Bailly C, Eastmond PJ, Penfield S (2017) Awake1, an ABC-Type Transporter, Reveals an Essential Role for Suberin in the Control of Seed Dormancy. Plant Physiol 174:276–283. <https://doi.org/10.1104/pp.16.01556>
- Francisco RM, Regalado A, Ageorges A, Burla BJ, Bassin B, Eisenach C, Zarrouk O, Vialet S, Marlin T, Chaves MM, Martinoia E, Nagy R (2013) ABCC1, an ATP binding cassette protein from grape berry, transports anthocyanidin 3-O-Glucosides. Plant Cell 25:1840–1854. <https://doi.org/10.1105/tpc.112.102152>
- Franke R, Schreiber L (2007) Suberin—a biopolyester forming apoplastic plant interfaces. Curr Opin Plant Biol 10:252–259. <https://doi.org/10.1016/j.pbi.2007.04.004>
- Fujita Y, Hara Y, Suga C, Morimoto T (1981) Production of shikonin derivatives by cell suspension cultures of *Lithospermum erythrorhizon*. Plant Cell Rep 1:61–63. <https://doi.org/10.1007/BF00269273>
- Fukui H, Yazaki K, Tataba M (1984) Two phenolic acids from *Lithospermum erythrorhizon* cell suspension cultures. Phytochemistry, 23: 2398–2399. [https://doi.org/10.1016/S0031-9422\(00\)80569-5](https://doi.org/10.1016/S0031-9422(00)80569-5)
- Garcia O, Bouige P, Forestier C, Dassa E (2004) Inventory and comparative analysis of rice and Arabidopsis ATP-binding cassette (ABC) systems. J Mol Biol 343:249–265. <https://doi.org/10.1016/j.jmb.2004.07.093>
- Geldner N, Dénervaud-Tendon V, Hyman DL, Mayer U, Stierhof YD, Chory J (2009) Rapid, combinatorial analysis of membrane compartments in intact plants with a multicolor marker set. Plant J 59:169–178. <https://doi.org/10.1111/j.1365-313X.2009.03851.x>
- Gräfe K, Schmitt L (2021) The ABC transporter G subfamily in *Arabidopsis thaliana*. J Exp Bot 72:92–106. <https://doi.org/10.1093/jxb/eraa260>
- Gräfe K, Shanmugarajah K, Zobel T, Weidtkamp-Peters S, Kleinschrodt D, Smits SHJ, Schmitt L

- (2019) Cloning and expression of selected ABC transporters from the *Arabidopsis thaliana* ABCG family in *Pichia pastoris*. PLOS ONE 14:e0211156. <https://doi.org/10.1371/journal.pone.0211156>
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95–98.
- Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcotili P, Nielsen H, Krogh A, Winther O (2022) DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. <https://www.biorxiv.org/content/10.1101/2022.04.08.487609v1.article-info> <https://doi.org/10.1101/2022.04.08.487609>
- He Y, Xu J, Wang X, He X, Wang Y, Zhou J, Zhang S, Meng X (2019) The Arabidopsis Pleiotropic Drug Resistance Transporters PEN3 and PDR12 Mediate Camalexin Secretion for Resistance to *Botrytis cinerea*. Plant Cell 31:2206–2222. <https://doi.org/10.1105/tpc.19.00239>
- Ichino T, Yazaki K (2022) Modes of secretion of plant lipophilic metabolites via ABCG transporter-dependent transport and vesicle-mediated trafficking. Curr Opin Plant Biol 66:102184. <https://doi.org/10.1016/j.pbi.2022.102184>
- Ito E, Munakata R, Yazaki K (2023) Gromwell, a purple link between traditional Japanese culture and plant science. Plant Cell Physiol., 64 :567-570. <https://doi.org/10.1093/pcp/pcad038>
- Izuishi Y, Isaka N, Li H, Nakanishi K, Kageyama J, Ishikawa K, Shimada T, Masuta C, Yoshikawa N, Kusano H, Yazaki K (2020) Apple latent spherical virus (ALS-V)-induced gene silencing in a medicinal plant, *Lithospermum erythrorhizon*. Sci Rep 10:13555. <https://doi.org/10.1038/s41598-020-70469-1>
- Kang J, Park J, Choi H, Burla B, Kretschmar T, Lee Y, Martinoia E (2011) Plant ABC Transporters. Arabidopsis Book 9:e0153.
- Kiyoto S, Ichino T, Awano T, Yazaki K (2022) Improved chemical fixation of lipid-secreting plant cells for transmission electron microscopy. Microscopy 71:206–213. <https://doi.org/10.1093/jmicro/dfac018>
- Kok JW, Klappe K, Hummel I (2014) The role of the actin cytoskeleton and lipid rafts in the localization and function of the ABCC1 transporter. Adv Biol 2014:105898. <https://doi.org/10.1155/2014/105898>
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Li FS, Phyto P, Jacobowitz J, Hong M, Weng JK (2019) The molecular structure of plant sporopollenin.

- Nat Plants 5:41–46. <https://doi.org/10.1038/s41477-018-0330-7>
- Li H, Matsuda H, Tsuboyama A, Munakata R, Sugiyama A, Yazaki K (2022) Inventory of ATP-binding cassette proteins in *Lithospermum erythrorhizon* as a model plant producing divergent secondary metabolites. DNA Res 29 :dsac016. <https://doi.org/10.1093/dnares/dsac016>
- Li M, Wang F, Li S, Yu G, Wang L, Li Q, Zhu X, Li Z, Yuan L, Liu P (2020) Importers Drive Leaf-to-Leaf Jasmonic Acid Transmission in Wound-Induced Systemic Immunity. Mol Plant 13:1485–1498. <https://doi.org/10.1016/j.molp.2020.08.017>
- Li Q, Zheng J, Li S, Huang G, Skilling SJ, Wang L, Li L, Li M, Yuan L, Liu P (2017) Transporter-Mediated Nuclear Entry of Jasmonoyl-Isoleucine Is Essential for Jasmonate Signaling. Mol Plant 10:695–708. <https://doi.org/10.1016/j.molp.2017.01.010>
- Linsmaier EM, Skoog F (1965) Organic growth factor requirements of tobacco tissue cultures. Physiol Plant 18:100–127. <https://doi.org/10.1111/j.1399-3054.1965.tb06874.x>
- Linton K.J (2007) Structure and function of ABC transporters. Physiology (Bethesda) 22:122–130. <https://doi.org/10.1152/physiol.00046.2006>
- Liu L, Zhao L, Chen P, Cai H, Hou Z, Jin X, Aslam M, Chai M, Lai L, He Q, Liu Y, Huang X, Chen H, Chen Y, Qin Y (2020) ATP binding cassette transporters ABCG1 and ABCG16 affect reproductive development via auxin signalling in Arabidopsis. Plant J 102:1172–1186. <https://doi.org/10.1111/tpj.14690>
- McFarlane HE, Shin JJH, Bird DA, Samuels AL (2010) Arabidopsis ABCG Transporters, Which Are Required for Export of Diverse Cuticular Lipids, Dimerize in Different Combinations. Plant Cell 22:3066–3075. <https://doi.org/10.1105/tpc.110.077974>
- Nakagawa T, Suzuki T, Murata S, Nakamura S, Hino T, Maeo K, Tabata R, Kawai T, Tanaka K, Niwa Y, Watanabe Y, Nakamura K, Kimura T, Ishiguro S (2007) Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. Biosci Biotechnol Biochem 71:2095–2100. <https://doi.org/10.1271/bbb.70216>
- Nakanishi K, Li H, Ichino T, Tatsumi K, Osakabe K, Watanabe B, Shimomura K, Yazaki K (2024) Peroxisomal 4-coumaroyl-CoA ligases participate in shikonin production in *Lithospermum erythrorhizon*. Plant Physiol 195:2843–2859. <https://doi.org/10.1093/plphys/kiac157>
- Nawrath C, Schreiber L, Franke RB, Geldner N, Reina-Pinto JJ, Kunst L (2013) Apoplastic diffusion barriers in Arabidopsis. Arabidopsis Book 11:e0167. <https://doi.org/10.1199/tab.0167>
- Ofori PA, Mizuno A, Suzuki M, Martinoia E, Reuscher S, Aoki K, Shibata D, Otagaki S, Matsumoto S, Shiratake K (2018) Genome-wide analysis of ATP binding cassette (ABC) transporters in tomato. PLOS ONE 13:e0200854. <https://doi.org/10.1371/journal.pone.0200854>
- Ohara K, Muroya A, Fukushima N, Yazaki K (2009) Functional characterization of LePGT1, a membrane-bound prenyltransferase involved in the geranylation of *p*-hydroxybenzoic acid. Biochem J 421:231–241. <https://doi.org/10.1042/BJ20081968>

- 1
- 2
- 3 828 Oshikiri H, Li H, Manabe M, Yamamoto H, Yazaki K, Takanashi K (2024) Comparative Analysis of
- 4 829 Shikonin and Alkannin Acyltransferases Reveals Their Functional Conservation in
- 5 830 Boraginaceae. *Plant Cell Physiol* 65:362–371. <https://doi.org/10.1093/pcp/pcad158>
- 7 831 Oshikiri H, Watanabe B, Yamamoto H, Yazaki K, Takanashi K (2020) Two BAHD Acyltransferases
- 8 832 Catalyze the Last Step in the Shikonin/Alkannin Biosynthetic Pathway. *Plant Physiol* 184:753–
- 10 833 761. <https://doi.org/10.1104/pp.20.00207>
- 12 834 Panikashvili D, Savaldi-Goldstein S, Mandel T, Yifhar T, Franke RB, Höfer R, Schreiber L, Chory J,
- 13 835 Aharoni A (2007) The Arabidopsis *DESPERADO/AtWBC11* transporter is required for cutin
- 14 836 and wax secretion. *Plant Physiol* 145:1345–1460. <https://doi.org/10.1104/pp.107.105676>
- 17 837 Panikashvili D, Shi JX, Schreiber L, Aharoni A (2011) The Arabidopsis ABCG13 transporter is
- 18 838 required for flower cuticle secretion and patterning of the petal epidermis. *New Phytol*
- 19 839 190:113–124. <https://doi.org/10.1111/j.1469-8137.2010.03608.x>
- 21 840 Pichersky E, Lewinsohn E (2011) Convergent evolution in plant specialized metabolism. *Annu Rev*
- 22 841 *Plant Biol* 62:549–566. <https://doi.org/10.1146/annurev-arplant-042110-103814>
- 24 842 Pighin JA, Zheng H, Balakshin LJ, Goodman IP, Western TL, Jetter R, Kunst L, Samuels AL (2004)
- 25 843 Plant Cuticular Lipid Export Requires an ABC Transporter. *Science*, 306:702–704.
- 27 844 <https://doi.org/10.1126/science.1102331>
- 29 845 Pollard M, Beisson F, Li Y, Ohlrogge JB (2008) Building lipid barriers: biosynthesis of cutin and
- 30 846 suberin. *Trends Plant Sci* 13:236–246. <https://doi.org/10.1016/j.tplants.2008.03.003>
- 32 847 Quilichini TD, Samuels AL, Douglas CJ (2014) ABCG26-mediated polyketide trafficking and
- 33 848 hydroxycinnamoyl spermidines contribute to pollen wall exine formation in *Arabidopsis*. *Plant*
- 35 849 *Cell* 26 :4483–4498. <https://doi.org/10.1105/tpc.114.130484>
- 37 850 Rai A, Nakaya T, Shimizu Y, Rai M, Nakamura M, Suzuki H, Saito K, Yamazaki M (2018) *De Novo*
- 38 851 Transcriptome Assembly and Characterization of *Lithospermum officinale* to Discover Putative
- 39 852 Genes Involved in Specialized Metabolites Biosynthesis. *Planta Medica* 84:920–934.
- 41 853 <https://doi.org/10.1055/a-0630-5925>
- 43 854 Samuels L, Kunst L, Jetter R (2008) Sealing plant surfaces: cuticular wax formation by epidermal
- 45 855 cells. *Annu Rev Plant Biol* 59:683–707.
- 47 856 <https://doi.org/10.1146/annurev.arplant.59.103006.093219>
- 48 857 Segami S, Makino S, Miyake A, Asaoka M, Maeshima M (2014) Dynamics of vacuoles and H⁺-
- 50 858 pyrophosphatase visualized by monomeric green fluorescent protein in *Arabidopsis*: artifactual
- 51 859 bulbs and native intravacuolar spherical structures. *Plant Cell* 26:3416–3434.
- 53 860 <https://doi.org/10.1105/tpc.114.127571>
- 54 861 Shanmugarajah K, Linka N, Gräfe K, Smits SHJ, Weber APM, Zeier J, Schmitt L (2019) ABCG1
- 56 862 contributes to suberin formation in *Arabidopsis thaliana* roots. *Scientific Reports*, 9:11381.
- 58 863 <https://doi.org/10.1038/s41598-019-47916-9>
- 59
- 60
- 61
- 62
- 63
- 64
- 65

- Shibata Y, Ojika M, Sugiyama A, Yazaki K, Jones DA, Kawakita K, Takemoto D (2016) The Full-Size ABCG Transporters Nb-ABCG1 and Nb-ABCG2 Function in Pre- and Postinvasion Defense against *Phytophthora infestans* in *Nicotiana benthamiana*. *Plant Cell* 28:1163-1181. <https://doi.org/10.1105/tpc.15.00721>
- Song W, Zhuang Y, Liu T (2020) Potential role of two cytochrome P450s obtained from *Lithospermum erythrorhizon* in catalyzing the oxidation of geranylhydroquinone during Shikonin biosynthesis. *Phytochemistry* 175:112375. <https://doi.org/10.1016/j.phytochem.2020.112375>
- Song W, Zhuang Y, Liu T (2021) CYP82AR Subfamily Proteins Catalyze C-1' Hydroxylations of Deoxyshikonin in the Biosynthesis of Shikonin and Alkannin. *Org Lett* 23:2455–2459. <https://doi.org/10.1021/acs.orglett.1c00360>
- Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings of the 6th International Conference on Intelligent Systems for Molecular Biology* 6:175–182.
- Stecher G, Tamura K, Kumar S (2020) Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol Biol Evol* 37:1237–1239. <https://doi.org/10.1093/molbev/msz312>
- Suttiyut T, Auber RP, Ghaste M, Kane CN, McAdam SAM, Wisecaver JH, Widhalm JR (2022) Integrative analysis of the shikonin metabolic network identifies new gene connections and reveals evolutionary insight into shikonin biosynthesis. *Horticul Res* 9:uhab087. <https://doi.org/10.1093/hr/uhab087>
- Takanashi K, Nakagawa Y, Aburaya S, Kaminade K, Aoki W, Saida-Munakata Y, Sugiyama A, Ueda M, Yazaki K (2019) Comparative proteomic analysis of *Lithospermum erythrorhizon* reveals regulation of a variety of metabolic enzymes leading to comprehensive understanding of the shikonin biosynthetic pathway. *Plant Cell Physiol* 60:19–28. <https://doi.org/10.1093/pcp/pcy183>
- Tang C (2021) Exploring the evolutionary process of alkannin/shikonin *O*-acyltransferases by a reliable *Lithospermum erythrorhizon* genome. *DNA Res* 28:dsab015. <https://doi.org/10.1093/dnares/dsab015>
- Tang CY, Li S, Wang YT, Wang X (2020) Comparative genome/transcriptome analysis probes Boraginales' phylogenetic position, WGDs in Boraginales, and key enzyme genes in the alkannin/shikonin core pathway. *Mol Ecol Res* 20:228–241. <https://doi.org/10.1111/1755-0998.13104>
- Tang YJ, Wang Q, Xue JY, Li Y, Li RM, Nocker SV, Wang YJ, Zhang CH (2018) Gene cloning and expression analyses of *WBC* genes in the developing grapevine seeds. *J Integ Agri* 17:1348–1359. [https://doi.org/10.1016/S2095-3119\(17\)61827-6](https://doi.org/10.1016/S2095-3119(17)61827-6)
- Tatsumi K, Ichino T, Isaka N, Sugiyama A, Moriyoshi E, Okazaki Y, Higashi Y, Kajikawa M, Tsuji

Y, Fukuzawa H, Toyooka K, Sato M, Ichi I, Shimomura K, Ohta H, Saito K, Yazaki K (2023) Excretion of triacylglycerol as a matrix lipid facilitating apoplastic accumulation of a lipophilic metabolite shikonin. *J Exp Bot* 74:104–117. <https://doi.org/10.1093/jxb/erac405>

Tatsumi K, Ichino T, Onishi N, Shimomura K, Yazaki K (2020) Highly efficient method of *Lithospermum erythrorhizon* transformation using domestic *Rhizobium rhizogenes* strain A13. *Plant Biotechnol (Tokyo)* 37:39–46. <https://doi.org/10.5511/plantbiotechnology.19.1212a>

Tatsumi K, Yano M, Kaminade K, Sugiyama A, Sato M, Toyooka K, Aoyama T, Sato F, Yazaki K (2016) Characterization of Shikonin Derivative Secretion in *Lithospermum erythrorhizon* Hairy Roots as a Model of Lipid-Soluble Metabolite Secretion from Plants. *Front Plant Sci* 7:1066. <https://doi.org/10.3389/fpls.2016.01066>

Uemura T, Yoshimura SH, Takeyasu K, Sato MH (2002) Vacuolar membrane dynamics revealed by GFP-AtVam3 fusion protein. *Genes Cells* 7 :743–753. <https://doi.org/10.1046/j.1365-2443.2002.00550.x>

Ueoka H, Sasaki K, Miyawaki T, Ichino T, Tatsumi K, Suzuki S, Yamamoto H, Sakurai N, Suzuki H, Shibata D, Yazaki K (2020) A Cytosol-Localized Geranyl Diphosphate Synthase from *Lithospermum erythrorhizon* and Its Molecular Evolution. *Plant Physiol* 182:1933–1945. <https://doi.org/10.1104/pp.19.00999>

Verrier PJ, Bird D, Burla B, Dassa E, Forestier C, Geisler M, Klein M, Kolukisaoglu U, Lee Y, Martinoia E, Murphy A, Rea PA, Samuels L, Schulz B, Spalding EJ, Yazaki K, Theodoulou FL (2008) Plant ABC proteins--a unified nomenclature and updated inventory. *Trends Plant Sci* 13:151–159. <https://doi.org/10.1016/j.tplants.2008.02.001>

Wang L, Yao W, Zhang X, Tang Y, Van Nocker S, Wang Y, Zhang C (2023) The putative ABCG transporter *VviABCG20* from grapevine (*Vitis vinifera*) is strongly expressed in the seed coat of developing seeds and may participate in suberin biosynthesis. *Physiol Mol Biol Plants* 29:23–34. <https://doi.org/10.1007/s12298-022-01276-3>

Wang S, Wang R, Liu T, Lv C, Liang J, Kang C, Zhou L, Guo J, Cui G, Zhang Y, Werck-Reichhart D, Guo L, Huang L (2019) CYP76B74 Catalyzes the 3"-Hydroxylation of Geranyhydroquinone in Shikonin Biosynthesis. *Plant Physiol* 179:402–414. <https://doi.org/10.1104/pp.18.01056>

Wang X, He Z, Yang H, He C, Wang C, Fazal A, Lai X, Yang L, Wen Z, Yang M, Ma S, Jie W, Cai J, Yin T, Liu B, Yang Y, Qi J (2022) Genome-Wide Identification of *LeBAHDs* in *Lithospermum erythrorhizon* and In Vivo Transgenic Studies Confirm the Critical Roles of *LeBAHD1/LeSAT1* in the Conversion of Shikonin to Acetylshikonin. *Life (Basel)* 12:1775. <https://doi.org/10.3390/life12111775>

Wu FY, Tang CY, Guo YM, Bian ZW, Fu JY, Lu GH, Qi JL, Pang YJ, Yang YH (2017) Transcriptome analysis explores genes related to shikonin biosynthesis in *Lithospermeae*

plants and provides insights into Boraginales' evolutionary history. *Sci Rep* 7:4477.
<https://doi.org/10.1038/s41598-017-04750-1>

Yadav S, Sharma A, Nayik GA, Cooper R, Bhardwaj G, Sohal HS, Mutreja V, Kaur R, Areche FO, AlOudat M, Shaikh AM, Kovács B, Mohamed Ahmed AE (2022) Review of Shikonin and Derivatives: Isolation, Chemistry, Biosynthesis, Pharmacology and Toxicology. *Front Pharmacol* 13:905755. <https://doi.org/10.3389/fphar.2022.905755>

Yadav V, Molina I, Ranathunge K, Castillo IQ, Rothstein SJ, Reed, JW (2014) ABCG Transporters Are Required for Suberin and Pollen Wall Extracellular Barriers in *Arabidopsis*. *Plant Cell* 26:3569–3588. <https://doi.org/10.1105/tpc.114.129049>

Yan C, Duan W, Lyu S, Li Y, Hou, X (2017) Genome-Wide Identification, Evolution, and Expression Analysis of the ATP-Binding Cassette Transporter Gene Family in *Brassica rapa*. *Front Plant Sci* 8:349. <https://doi.org/10.3389/fpls.2017.00349>

Yamamoto H, Tsukahara M, Yamano Y, Wada A, Yazaki K (2020) Alcohol Dehydrogenase Activity Converts 3"-Hydroxy-geranylhydroquinone to an Aldehyde Intermediate for Shikonin and Benzoquinone Derivatives in *Lithospermum erythrorhizon*. *Plant Cell Physiol* 61:1798–1806. <https://doi.org/10.1093/pcp/pcaa108>

Yamamura Y, Sahin FP, Nagatsu A, Mizukami H (2003) Molecular cloning and characterization of a cDNA encoding a novel apoplastic protein preferentially expressed in a shikonin-producing callus strain of *Lithospermum erythrorhizon*. *Plant Cell Physiol* 44:437–446. <https://doi.org/10.1093/pcp/pcg057>

Yazaki K (2017) *Lithospermum erythrorhizon* cell cultures: Present and future aspects. *Plant Biotechnol (Tokyo)* 34:131–142. <https://doi.org/10.5511/plantbiotechnology.17.0823a>

Yazaki, K, Arimura, G, Ohnishi, T (2017) "Hidden" terpenoids in plants: Their biosynthesis, localisation and ecological roles. *Plant Cell Physiol* 58:1615-1621. doi: 10.1093/pcp/pcx123.

Yazaki K, Kuniyoshi M, Fujisaki T, Sato F (2002) Geranyl diphosphate:4-hydroxybenzoate geranyltransferase from *Lithospermum erythrorhizon*. Cloning and characterization of a key enzyme in shikonin biosynthesis. *J Biol Chem* 277:6240–6246. <https://doi.org/10.1074/jbc.M106387200>

Zhang Y, Kilambi HV, Liu J, Bar H, Lazary S, Egbaria A, Ripper D, Charrier L, Belew ZM, Wulff N, Damodaran S, Nour-Eldin HH, Aharoni A, Ragni L, Strader L, Sade N, Weinstain R, Geisler M, Shani E (2021) ABA homeostasis and long-distance translocation are redundantly regulated by ABCG ABA importers. *Sci Adv* 7:eabf6069. <https://doi.org/10.1126/sciadv.abf6069>

Zhong Y, Xun W, Wang X, Tian S, Zhang Y, Li D, Zhou Y, Qin Y, Zhang B, Zhao G, Cheng X, Liu Y, Chen H, Li L, Osbourn A, Lucas WJ, Huang S, Ma Y, Shang Y (2022) Root-secreted bitter triterpene modulates the rhizosphere microbiota to improve plant fitness. *Nat Plants* 8:887–

896. <https://doi.org/10.1038/s41477-022-01201-2>

Zhou Y, Ma Y, Zeng J, Duan L, Xue X, Wang H, Lin T, Liu Z, Zeng K, Zhong Y, Zhang S, Hu Q, Liu M, Zhang H, Reed J, Moses T, Liu X, Huang P, Qing Z, Liu X, Tu P, Kuang H, Zhang Z, Osbourn A, Ro DK, Shang Y, Huang S (2016) Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat Plants* 2:16183. <https://doi.org/10.1038/nplants.2016.183>

Zhu Y, Chu SJ, Luo YL, Fu JY, Tang CY, Lu GH, Pang YJ, Wang XM, Yang RW, Qi JL, Yang YH (2018) Involvement of LeMRP, an ATP-binding cassette transporter, in shikonin transport and biosynthesis in *Lithospermum erythrorhizon*. *Plant Biol (Stuttg.)* 20:365–373. <https://doi.org/10.1111/plb.12666>

Zhu Y, Lu GH, Bian ZW, Wu FY, Pang YJ, Wang XM, Yang RW, Tang CY, Qi JL, Yang YH (2017) Involvement of *LeMDR*, an ATP-binding cassette protein gene, in shikonin transport and biosynthesis in *Lithospermum erythrorhizon*. *BMC Plant Biol* 17:198. <https://doi.org/10.1186/s12870-017-1148-6>

Zuo J, Niu QW, Chua NH (2000) Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *Plant J* 24:265–273. <https://doi.org/10.1046/j.1365-3113x.2000.00868.x>

Figure legends

Fig. 1 Phylogenetic analysis of LeABCG1 and LeABCG2 (a) Phylogenetic tree of half- and full-size ABCG transporter proteins from *L. erythrorhizon*, Arabidopsis, tomato, and PhABCG1. The polypeptide sequences of the conserved NBD domains (i.e., the region between the Walker A and Walker B positions) of ABCG proteins were used for the alignment. The maximum likelihood method (i.e., the JTT model) was used to construct the tree. Bootstrapping was performed with 1,000 replicates and the percentage of trees in which the associated taxa clustered together is shown next to the branches. Classes (A-G) and clades (I-III) as defined by Zhang et al. (2021) and Panikashvili et al. (2011) were used for classification. (b, c) Sequence alignment of the NBD domain. The NBD domains of LeABCG1 (b) and LeABCG2 (c) were compared with their corresponding orthologs. Conserved motifs in the NBD domain are indicated. All amino acids except for alanine and glycine are colored according to the Jmol color code (<http://jmol.sourceforge.net/jscolors/>).

Fig. 2 Genomic structures and motif locations of LeABCG1 and LeABCG2 (a) Genomic structures of *LeABCG1* and its Arabidopsis orthologs, *AtABCG2* and *AtABCG20*, and its tomato orthologs, *SlABCG8*, *SlABCG9*, and *SlABCG13*. (b) Genomic structures of *LeABCG2* and its Arabidopsis ortholog *AtABCG3* and its tomato orthologs, *SlABCG17* and *SlABCG28*. (c, d) The motif positions of the NBD domain in a and b are enlarged in c and d, respectively. Also shown are the genomic

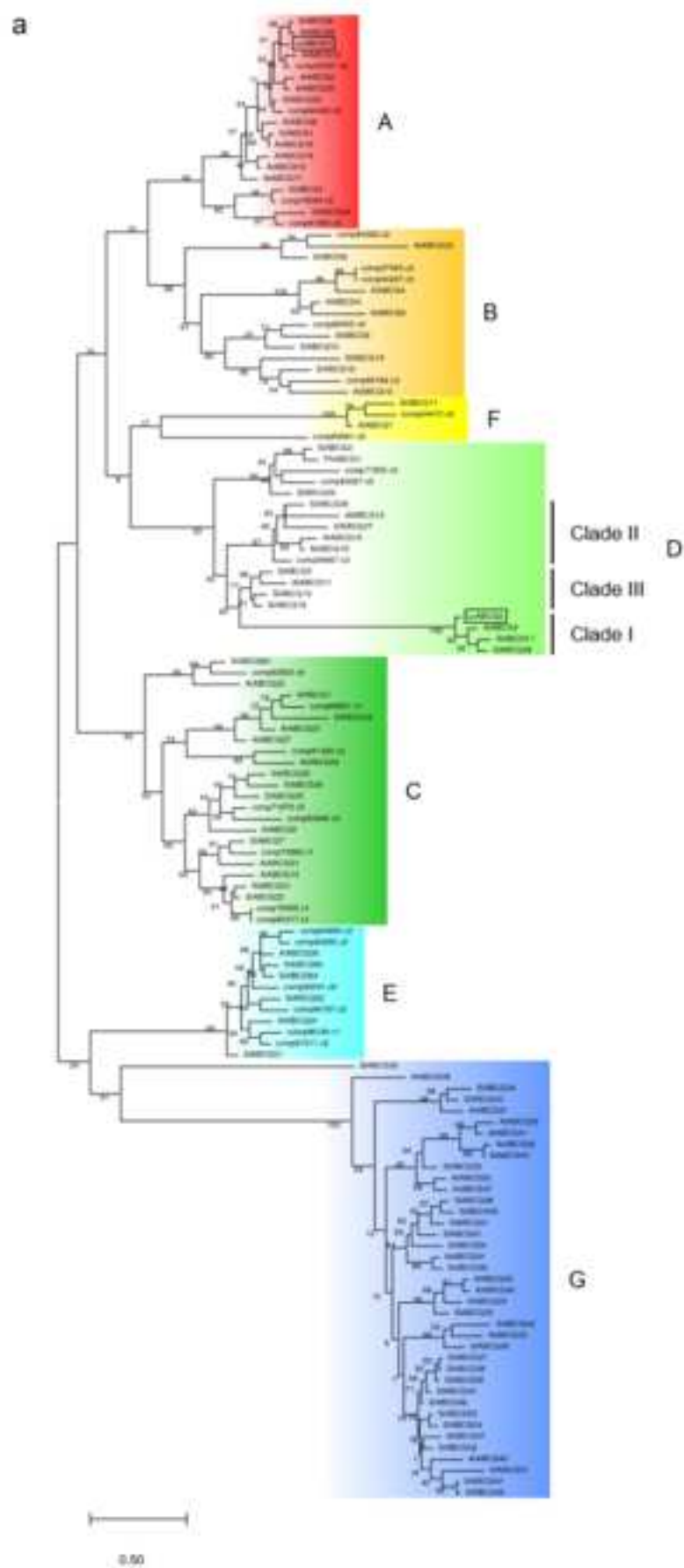
fragments between 300 nt and 1,100 nt distant from the start codon of the *LeABCG1* orthologs (**c**) and the third exon of the ORF region of the *LeABCG2* orthologs (**d**). Open boxes, closed boxes, and lines represent untranslated regions (UTRs), exons, and introns, respectively. In the translated polypeptide sequences, the following motifs are shown in colored boxes representing specific exons: Walker A (blue), Q-loop (orange), ABC signature (green), Walker B (purple), D-loop (red), potential H-loop (white), and transmembrane helix (TM; yellow).

Fig. 3 Gene expression of *LeABCG1* and *LeABCG2* in *L. erythrorhizon* (a, b) FPKM values of comp46786_c2_seq1 (**a**) and comp78121_c0_seq10 (**b**) from transcriptome data reported by Takanashi et al. (2019). Root bark, hairy roots cultured in the dark, and cells cultured in the dark produce shikonin (red bars). Contigs comp46786_c2_seq1 and comp78121_c0_seq10 correspond to *LeABCG1* and *LeABCG2*, respectively. (**c-h**) Gene expression of *LeABCG1* (**c-e**) and *LeABCG2* (**f-h**) analyzed by RT-qPCR. In the cultured cells (**c, f**) and the hairy roots (**d, g**), the presence (+) or absence (-) of secretion of shikonin derivatives and triacylglycerol (TAG) are shown at the bottom (n.d., not determined). *L. erythrorhizon* plants produce shikonin derivatives specifically in the root bark (**e, h**). Red bars represent shikonin-producing conditions. Each bar represents the mean with standard error of three independent cultures of the cells (**c, f**), three genetically independent lines of the hairy roots (**d, g**), and three plant individuals (**e, h**). *LeABCG* expression is normalized to *LeACT* gene expression. We set the expressions of LS dark in **c** and **f**, 1/2 MS medium in **d** and **g**, and peeled roots in **e** and **h** to 1.0 using the $\Delta\Delta CT$ method. The right panel in **e** is magnified from a part of the left panel. Statistically significant differences are not detected by Tukey HSD multiple comparison after one-way ANOVA (**c, e, f, h**) and Student's t-test (**d, g**).

Fig. 4 Gene expression of *LeABCG1* and *LeABCG2* in metabolite-treated *L. erythrorhizon* (a, b) Gene expression of *LeABCG1* (**a**) and *LeABCG2* (**b**) in the hairy roots treated with intracellular (red) or extracellular metabolite mixture (blue) prepared from *L. erythrorhizon* cells at a concentration of 3 mg/mL or 0.3 mg/mL, respectively. Samples were collected at 0, 4, 10, 24, and 72 hours. The bottom panel in (**a**) is an enlargement of the top panel. (**c, d**) Gene expression of *LeABCG1* (**c**) and *LeABCG2* (**d**) in the hairy roots treated for 72 hours with individual metabolites, which are secreted by *L. erythrorhizon* cells and monomers of plant apoplastic biopolymers. As a representative shikonin derivative 1 mM acetylshikonin, as a suberin monomer 1 mM 1-hexacosanol (C26:0 fatty alcohol), as cutin precursors 1 mM 16-hydroxyhexadecanoic acid (C16:0 ω -hydroxy fatty acid [ω -OH FA]) and 1 mM 2-monopalmitin (monoacylglycerol [MAG]), and 0.1 mM 1,2-palmitin-3-palmitolein (triacylglycerol [TAG]) were used for the treatment. Each bar graph represents the mean with standard error of four or five independent treatments (**a, b**) and seven independent treatments (**c, d**). *LeABCG* expression is normalized to *LeACT* gene expression. We set the expressions of the sham treatment

(dimethyl sulfoxide) to 1.0 using the $\Delta\Delta\text{CT}$ method. Statistically significant differences at the same time point (**b**) and between different treatments (**d**) are indicated by different letters using Tukey HSD multiple comparison ($p < 0.05$) after one-way ANOVA; insignificant differences at the same time point (**a, b**) and between different treatments (**c**) are shown without any letters.

Fig. 5 Subcellular localization of LeABCG1 and LeABCG2 (a) GFP fluorescence of LeABCG1-mGFP, mGFP-LeABCG1, LeABCG2-mGFP, and mGFP-LeABCG2 in transiently expressed *N. benthamiana* leaves after agroinfiltration. Green, GFP fluorescence; magenta, autofluorescence of chlorophyll in plastids. Similar signal patterns were observed in at least 60 epidermal cells (representing at least two independent experiments and a total of four *N. benthamiana* leaves in total) for each construct. Bars indicate 50 μm in the upper panels and 10 μm in the lower panels. (b, c) Co-expression of LeABCG1-mGFP (**b**) or LeABCG2-mGFP (**c**) with mRFP-AtPIP1;4 (top panels) or mRFP-AtVAM3 (bottom panels) in transiently expressed *N. benthamiana* leaves after agroinfiltration. AtPIP1;4 and AtVAM3 are membrane markers for PM-resident membrane protein and tonoplast-resident membrane protein, respectively. Both are derived from *Arabidopsis thaliana*. Green, GFP fluorescence; Magenta, RFP fluorescence. Similar signal patterns were observed in at least 30 epidermal cells (representing three independent experiments and a total of six *N. benthamiana* leaves in total) for each combination. Bars indicate 20 μm in the left panels and 10 μm in the three right panels.

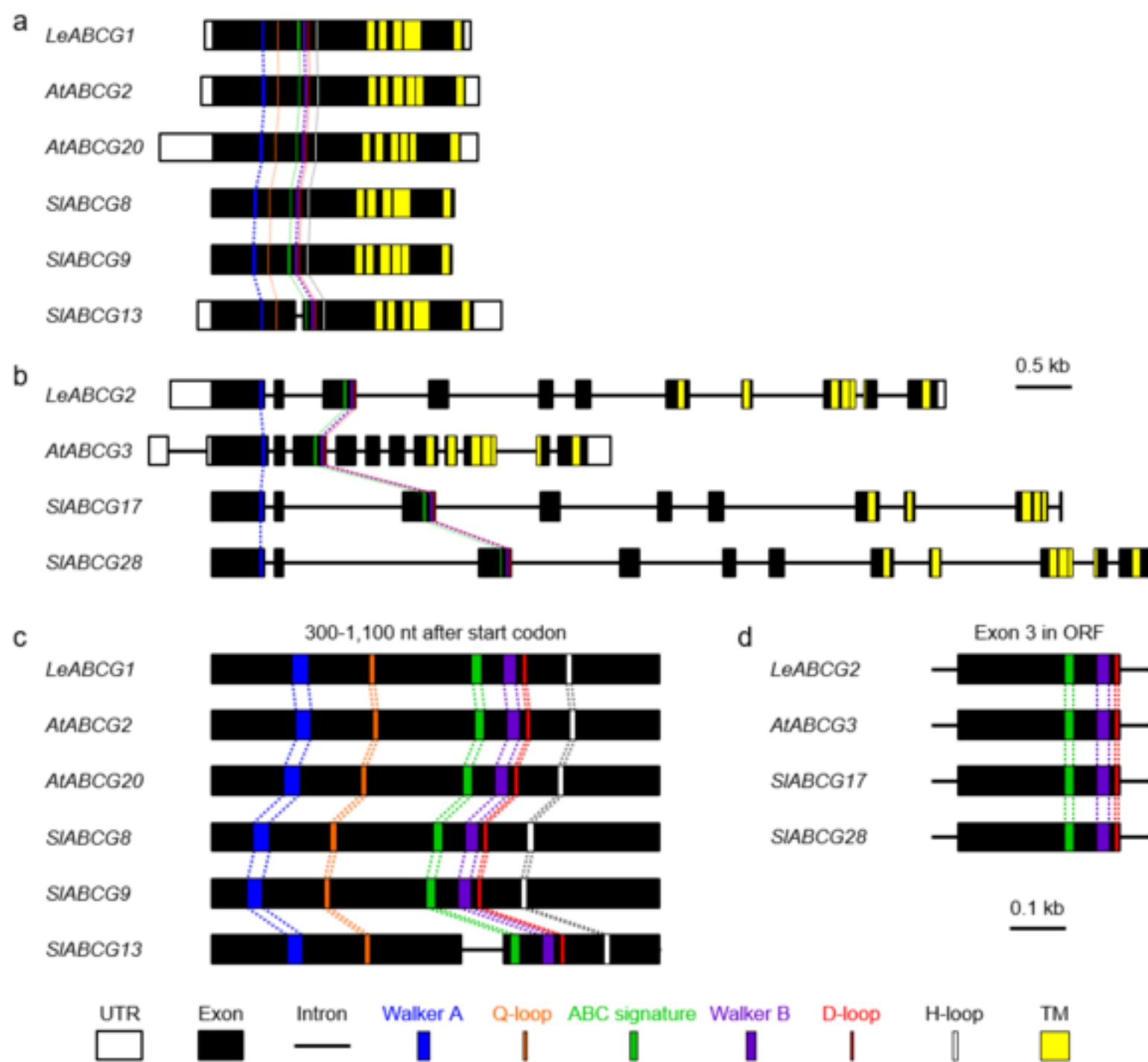


b

LeABCG1	149	GA	<u>SGSGKSTL</u>	IDALANRIS	RGSLKGSITL	NGEQL	ESKLLKVISAYV	<u>QDD</u>	198	
AtABCG2	151	GA	<u>SGSGKSTL</u>	IDALANRI	AKDSL	RGSITL	NGEVLESS	MOKVISAYV	<u>QDD</u>	200
AtABCG20	144	GA	<u>SGSGKSTL</u>	IDALANRIS	KESL	RGDITL	NGEVLESS	LHKVISAYV	<u>QDD</u>	193
SIABCG8	126	GA	<u>SGSGKSTL</u>	IDALANRIS	KDSL	KGEHKL	NGEPLHS	KLLKVISAYV	<u>QDD</u>	175
SIABCG9	122	GA	<u>SGSGKSTL</u>	IDALANRI	AKESL	KGEHKL	NGEKLHT	KLLKVISAYV	<u>QDD</u>	171
SIABCG13	146	GA	<u>SGSGKSTL</u>	IDALANRI	AKESL	KGFTITL	NGEPLDS	RLLKVISAYV	<u>QDD</u>	195
Walker A				Q-loop						
LeABCG1	199	LLYPMLTVE	ETLNFSAEFRLP	PSLSKSKKKS	RV	EALIDQLGLR	NAAKTII	248		
AtABCG2	201	LLFPMLTVE	ETLNFSAEFRLP	RSLSKSKKKA	RV	QALIDQLGLR	SAAKTVI	250		
AtABCG20	194	LLFPMLTVE	ETLNFSAEFRLP	SSLSKSKKKA	RV	QALIDQLGLR	NAAKTVI	243		
SIABCG8	176	LLYPMLTVE	ETLNFSAEFRLP	RTLSKSKKKS	RV	QALIDQLGLR	NAAKTII	225		
SIABCG9	172	LLYPMLTVE	ETLNFSAEFRLP	RTLSKSKKKS	RV	QALIDQLGLR	NAAKTII	221		
SIABCG13	196	LLYPMLTVE	ETLNFSAEFRLP	RNLSNSKSKK	RV	QALIDQLGLR	NAAKTII	245		
LeABCG1	249	GDEGHRGVS	GGERRRVSIG	IDIIHDP	ILLFLDEPT	SGLDSTSAF	HWVKVL	298		
AtABCG2	251	GDEGHRGVS	GGERRRVSIG	NDIIHDP	IILFLDEPT	SGLDSTSA	YHWIKVL	300		
AtABCG20	244	GDEGHRGVS	GGERRRVSIG	TDIIHDP	IILFLDEPT	SGLDSTSA	YHWVKVL	293		
SIABCG8	226	GDEGHRGVS	GGERRRVSIG	IDIIHDP	IILFLDEPT	SGLDSTSAF	HWVKVL	275		
SIABCG9	222	GDEGHRGVS	GGERRRVSIG	IDIIHDP	IILFLDEPT	SGLDSTSAF	HWVKVL	271		
SIABCG13	246	GDEGHRGVS	GGERRRVSIG	IDIIHDP	IILFLDEPT	SGLDSTSA	YHWVKVL	295		
ABC signature				Walker B	D-loop					
LeABCG1	299	QRIAGSGSIV	INSIHQ						314	
AtABCG2	301	QRIAGSGSIV	INSIHQ						316	
AtABCG20	294	QRIAGSGSIV	INSIHQ						309	
SIABCG8	276	QRIAGSGSIV	INSIHQ						291	
SIABCG9	272	QRIAGSGSIV	INSIHQ						287	
SIABCG13	296	QRIAGSGSIV	INSIHQ						311	
H-loop?										

c

LeABCG2	143	GPAKSGKST	LLRALSGKL	PD	SAR	HYGEVF	INGSRSHLPYGSYG	YVERETT	192
AtABCG3	151	GPAKSGKST	LLRALAGRLP	PSAK	HYGEVF	VNGSKSH	WPYGSYG	FVERETT	200
SIABCG17	143	GPAKSGKST	LLRALAGRLP	DSTR	HYGEVF	VNGTRRH	WPYGSYG	YDRETT	192
SIABCG28	143	GPAKSGKST	LLRSLAGRLP	DSAR	HYGEVF	VNGTKRS	WPYGSYG	FYDRETT	192
Walker A									
LeABCG2	193	LIGSLTVRE	FLYYSALLQL	PGFF	FQKKS	VVEDATLANS	LGDYANKLIGGH	242	
AtABCG3	201	LIGSLTVRE	FLYYSALLQL	PGFL	FQKRS	VVEDAIOANS	LSLDYANKLIGGH	250	
SIABCG17	193	LIGSLTVRE	FLYYSALLQL	PGFF	QRRS	VVEDAIDANS	LGDYANKLIGGH	242	
SIABCG28	193	LIGSLTVRE	FLYYSALLQL	PGFL	QKRS	VVEDAIDANS	LGDYANKLIGGH	242	
LeABCG2	243	CYKGLSSG	ERRRVSIA	RELV	RRPHALF	IDEPLYHLD	279		
AtABCG3	251	CYKGLRSG	ERRRVSIA	RELV	RRPHILF	IDEPLYHLD	287		
SIABCG17	243	CYTKGLRSG	ERRRVSIA	RELV	RRPHILF	IDEPLYHLD	279		
SIABCG28	243	CYKGLRSG	ERRRVSIA	RELV	RRPHILF	IDEPLYRLD	279		
ABC signature				Walker B	D-loop				



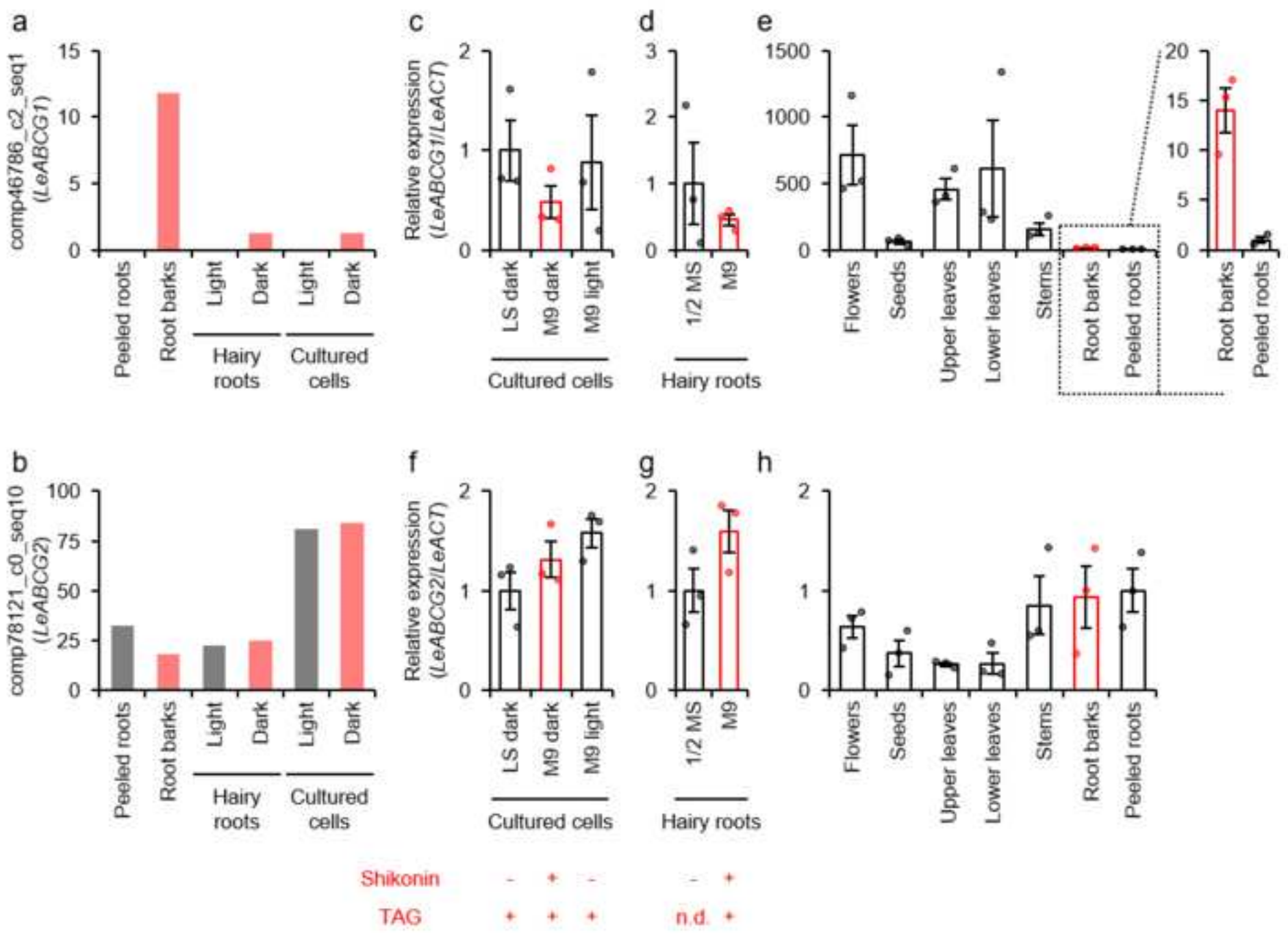
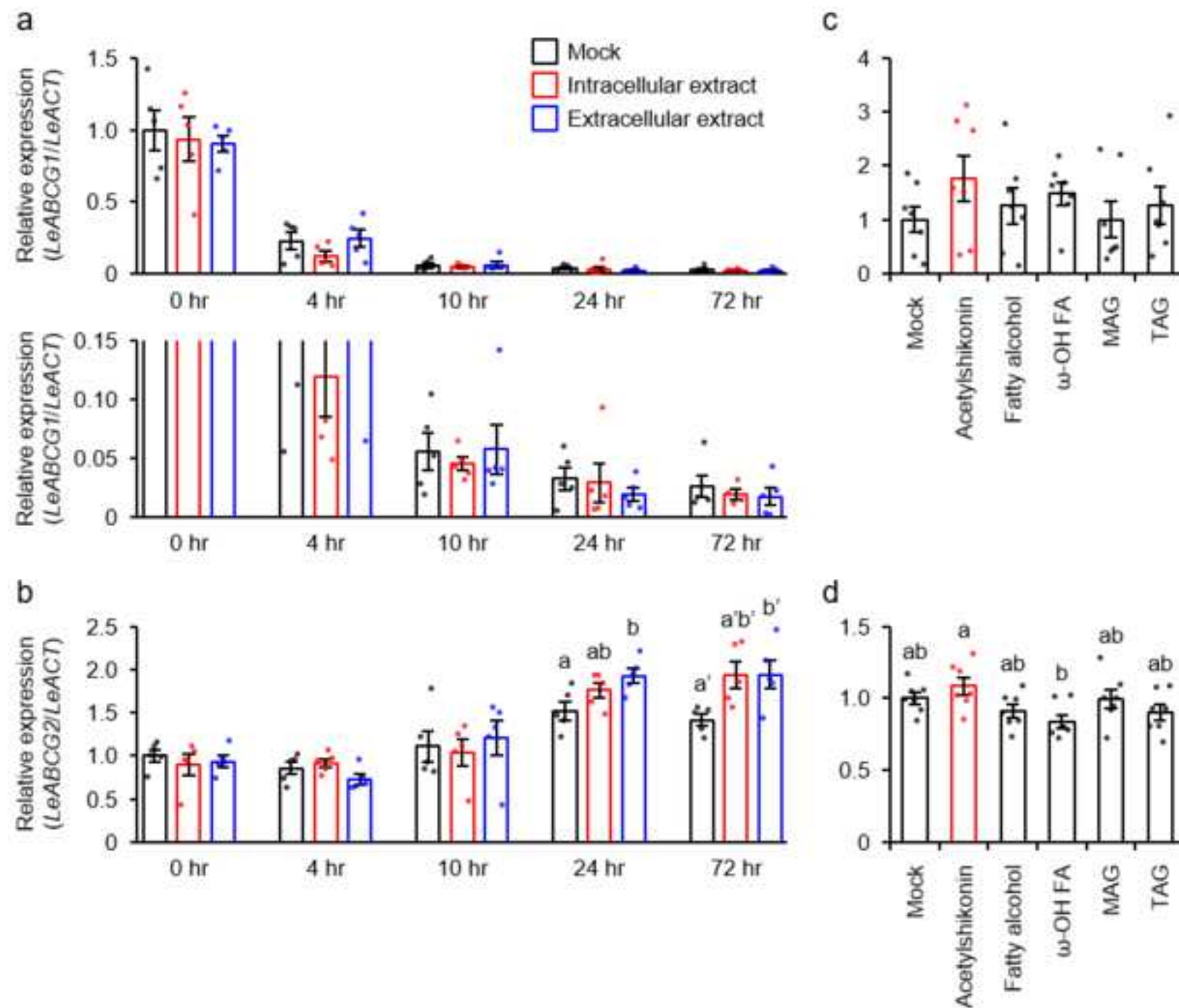
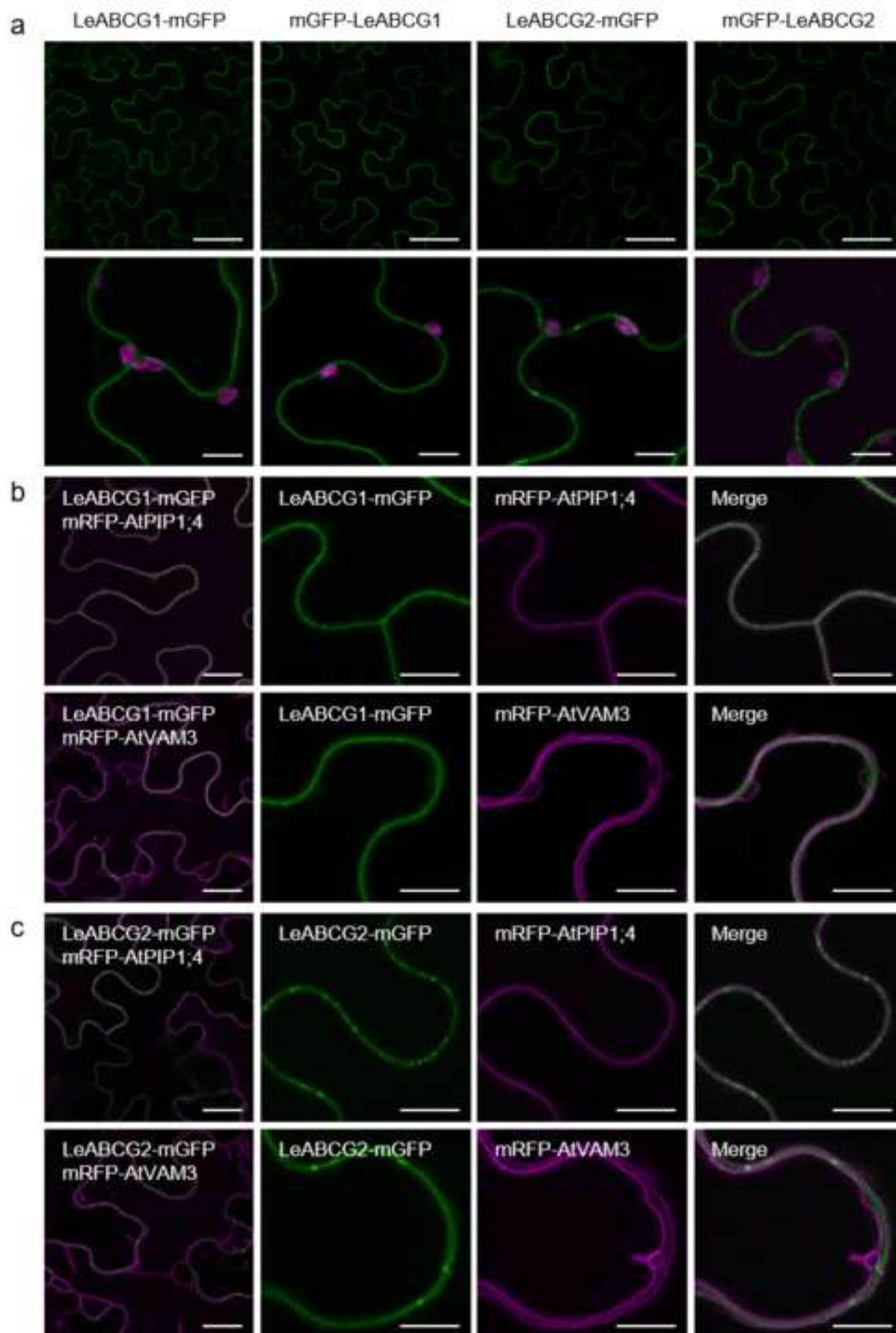
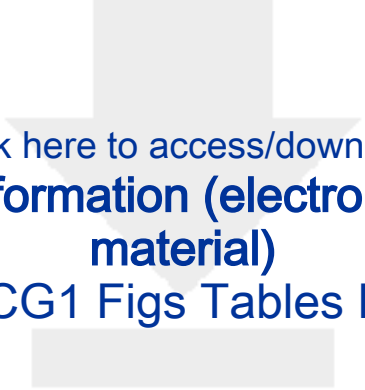


Fig. 4



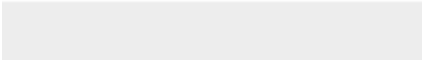
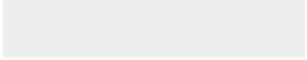




[Click here to access/download](#)

Supplementary Information (electronic supplementary material)

LeABCG1 Figs Tables R1.pdf



Supplementary Information

Title

Characterization of two G-type half-size ABC transporter genes from the lipid-secreting medicinal plant *Lithospermum erythrorhizon*

Journal

Journal of Plant Research

The names of the authors

Takuji Ichino^{1, 2}, Kanade Tatsumi^{1, 3}, Yuka Munakata¹, Ai Tsuboyama¹, Eiko Moriyoshi¹, Masaru Nakayasu¹, Kojiro Takanashi⁴, Koichiro Shimomura⁵, Kazufumi Yazaki^{1, *}

The affiliations of the authors

¹Research Institute for Sustainable Humanosphere, Kyoto University, Uji, Japan

²Laboratory of Medicinal Cell Biology, Kobe Pharmaceutical University, Kobe, Japan

³Institute of Plant Molecular Biology, Centre National de la Recherche Scientifique, University of Strasbourg, Strasbourg, France

⁴Department of Biology, Faculty of Science, Shinshu University, Matsumoto, Japan

⁵Graduate School of Life Sciences, Toyo University, Itakura, Japan

*Corresponding Author: Kazufumi Yazaki (yazaki@rish.kyoto-u.ac.jp)

>comp46786_c2_seq1 deduced polypeptide
VMQDDLPLYMLTVEETLMFSAEFRLLPSSLKSKKKSRVEALIDQLGLRNAAKTIIGDEGHRGVSGGERRRVSIGIDIHDPILLFLDEPTSG

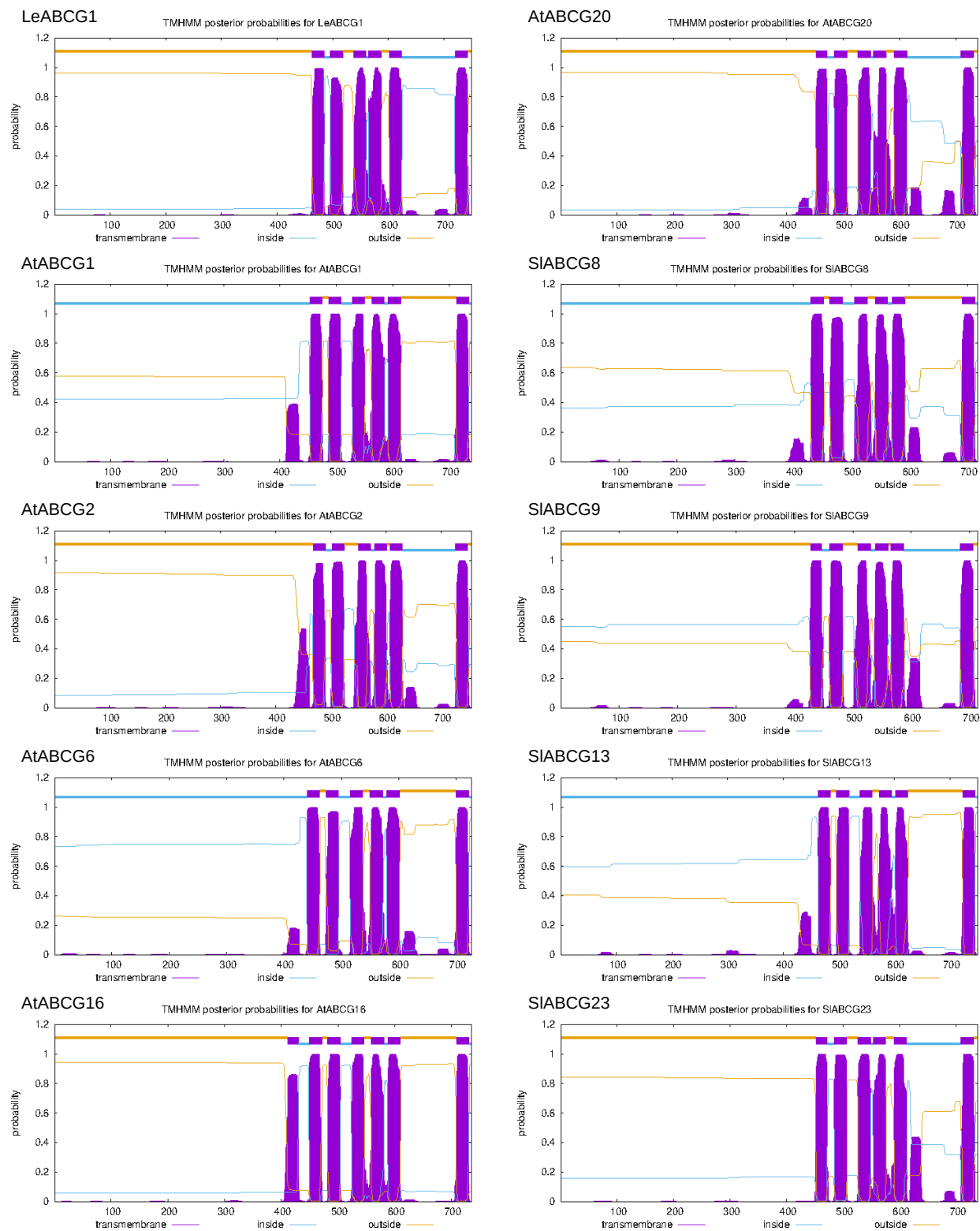
>comp78121_c0_seq10

CCTGATGATCCACACTTGGCCCTCATATTTACTTGGCTTTTGTGCTCTTATTTCTGCTTTATTCATGATGCTTCTGGATCCAACCAACCAACTTGTGTTTCTGTTATTTTATTTATTTA
TTTGGGTGCTTATTTAGTGTGTACTTACTTTTTTAAAGTTTGGTTCGAAATTTATTTATTTATTTAAAGTTCTTGGCAACTTGTTAACGTGTATAGTAAGGTTTATTTTATAGTTAA
AATTACAATAACTCCCGGCCCGTCCCTTATAATGAGGTTCGAAGCATTGAGCTCAAGATCAAGGGTATATGAAGTGTATAAAGTCTGAAACATGGTACTTTTTCTAGGGGGTTAGGTT
TATTAGATAAACTGTGTGCCATCTTAATTTTGTGTTTTTAAATTCCTGCAAGCTTGTGTGTATGACATGCTACAAAGTTGAAACTGAACTTTGTTAATTAGATAAAATG
TCTGCACTTTTTTACTAGTACTCATAGGTTATGTAGTTATTTAGCTTATTTGGGTAGCTGTTGTGTAGGTTTGGTAACTGGGGTTTTATTTTGAAGAACTTACTGGGGT
GTAAAGAGTTGAGCTTTAGGTGCACATGTAAATTTATGATGTTTTAGCTATCTAGATGAAGAAGCAACAACTTTGTACCTGAAATCTGATAAATCAGACCCCACTTGTTTT
GGTTATGTTTGCAAAGTTATTGGTATTGTTTGTGATTATATTTCTTAAAGTTCAAAGTTATAGTCATTTGTTTATGCATCTCATTGCAGAATTGTTTTGGATTTTACTTTTAAAG
ATGGAGGAAGTTTGAGTAGTTGAATTTGATTATGGAGCTCGAAGGAAATGAGGAGCAAGTTTCGGCTTTGTGGAATAGCTGTTTATAGCTAGATTGTTAGGCTAGATTGAAGATCATATGAGCAAGC
ATGGAGAGACATAACACATTCATGATCAATTAGGTCCTTCTTCACTTCAGCCAGCTAGCTCGCAAGTCGTTTACCCCTAGTAAATTTCTTCTACTCTAGGAAACCAAGAACGTTTAGG
AAACCCATTTCTTATGAAGATTACCCGGATTGGGAGGACAATGATGTTGAGGTAGGATGGAAGAAGGAGGTGATTCATCCATGTTGGCACCACCAACCGGATCTCCATCACTTTCGAAG
ATTACTAGTGGCTGATCGCATCTCTCCCTTGCCAACTGGTGCAAGTTGTGCGCAAGAAGATGCGCGGGGCTCAGTTGTTTGAAGGATTTAACAGTCACATATAAAGGGGAAGAGGAAG
TATCTGACAAGGCTTACAAGAGTCAAAGTGTACGCAATGCGCTGAGCAACATGACGGTAATAAATGGGCCCAAGCAAGTCAAGGAAATCAACTCTGTTGAGAGCTTCTTCAGGAAGAACT
CCTGATTCTGCGAAGATGTATGGAGAGGTGTTTATAAAGTGAGTGCAGATCAACTGCGCTATGAGTATGTTGAGAGGAGACTACATATTTGGTCTCTTGACAGTCGCT
GAATTCCTTTACTACTCAGCATGCTGTTCAAGTGCGCTGGATTTCTTTTTCAGAAAGAAGATGTTGATAGAGGATGCTATCCTTGCTATGCTCTGGGGGATATGCTAACCAACTCATAGGT
GGCCACTGCTACATGAAGGGTCTTTCAGTGGTGGAAGACAGGCGTGTAGCATTTGTAGCATAGTATGAGTATGAGACCTCATGCTCTTATTTAGACGAGGCTCTGTATCACTCTGACAGT
GTGCGTGCACTCTAATGATGGTGAGCTTGAAGAAGCTTGCAGACAGAGTGCACATCATATTTACCTATACCGAGAGCAAGCGAGTTTTGGCCCTTTTGACCGTATCTGCGCT
CTTTCAAATGGAACACATTTGTTCTTTGGAGAAACATTGGCTTGTGTCAGCACTTCTCAAATGCAAGGATTCCTTGTCCGATAATGCAAAGTCCATCAGATCACTTCTCTCGAGCGATA
AATACAGATTTGATGCGGATGCTGCTATGTGTAATAAATGGGAGGATGACATGCAAGATTATCATCAGTGAATAGGAGCAAGCTCAGTTGCAATACGCAACATGAAACCACTTACAAA
TCGTGACAGATGCTGCTGCTGTTGAAACCATGATGGAAGCTTCGACAAAGAGGGGTCCAAACTAATGAAGACGGGGAAGAGCTGATGAATGCAACACGAATTCGTGATTTAGCTTGG
AGCATCACTTATTAATCATGTGCAAGAGGTGGAAGTATTACTGGCTCGGGTAGTCTATATAGCTCTTGAGTCTGTTTGGTTGAGCATTTCCGGATTTGGGGGATTTCTGTGTCATCC
GTGGTGACCAGAGTTGCAGCAGCTCTCGTTTTCATTTCACTTTTTCAGCTCTTTTGAAGTATAGCCGCTGTACCTGCTCAGCTGAAGGAATCAAGATATTTGCGCTGTGAAGATCAATCATC
CATCTTGGGACATTTGGTTTTCTTCACTTGGACAACTTTTGGCCAGTATCCCCATCTTATTCCTCATCTTCCATCATCAAGTCTGTTTCTTATTTCTGTGTTGCTGCGAGATGAGTTC
AGCTTTGTGATGATCTTTGTCTGAAATTTCTTCATAGCGCTTCTCTCAAATGAAGTATGAGTCTGATGCTGATGCTGCTGGAATCAAACTTTCTGGACATCTCGTCAACTAATG
ATACAGCTTACGTTGATGCTTTCTGCTGGTATTTTGAATCAGAAGTGCTTTGCTGATGCTGGATGGAAGTATCCCATATCTTACATAGCTTTCCATACATACTCGATCCAGGGCCCT
CTGGAGAATGAGTATGTGGGAGCTTCTTTCGAGTTTGGACAGGTGAGGACATATCTGTTACAGTGTCTCTCGGAATGTATACGAAGATCTCTCGACAGATCAACTCCAAAGTGGAAAAAT
GTGCTAGTTATTTGCTCTCATGGCTGTGGGTGGCGCATCTTGGTTTGAATATTGTCGAGTTTCGTTGATGCAACACTGCTCGATCCGCAAACTGTCCGGTGCAATCAAGCAACAAC
AATTCAAGAATGATGATATACTCATGAAACCTTTTAGCTTTTGTAACTTTAAGCTTCCAACCTCATGTTGAATGGATGATATGATTTTATTTGTAATGATTCAACACTAGTTATTGCACTT
TATTTCTGCGGATACATGATGAGTGGCTATGTTGGCTACAGTAATATAGGAAGAGCTGCTTCTAGCTATGGCCAAATTTGGTTGGTATCGATCGGTTTTCTCGGGTAGGATACAGAGTAGT
TGGAGTACCTCTCTCGTTTATGGTACAGAGTGGTTTGGAGTACCTCTTACTGATTTTATGTGATACTCTATATTACTAATAAAAAATTAATTTATATGTGAAGAAAAA
AAAAA

>comp78121_c0_seq10 deduced polypeptide
MEEKIQSHSDHYRGLSSSSASSAPSLRPSNFFYLRLKGLPFRKPISEYSDPDWEDNVEVRMEEGGDISIVHAPTTDSPSLSKITSASLPSPLPTGAVVARKIAGASVVMVKDLTVTIKGRKR
YSDYIKSSNGYSGPTMTVTIMGPAKSGKSTLLRALSGKPLSDSARMYGEVFINGRSHLPYGSYGIVERTETLIGSLTVREFLYYSAALLPGFFGOKKSVVEDALMSLGDYANKLIG
GHYCMKGLSSGERRRRYSARELVMRPHALFIDEPLYHLDVSAALLMMVTLKKLASTGCTLFTIYQSSSTVEVGLFDRICLLNGNLTFFGETLACLOHFSNAGFPCCIMQSPSDHFLRAI
NTMDFRDIAMCKNWQDDHADLSVNNMDTAIRTLETITYKSSADAAAVETVILKAEKEGPTIKSRGSGNARIATVLRSLMLSRWEKYYWLRLLVYMLLSVCIGISAGFSGHLSLS
VTVRAAVFVIFSAALLSIAAVPALDKEIKIFACEEENSHSGTLVGLQGLSSIPFLFLISISSLVYFVLGDRDEFSLMLYFVNFCTKLNLNEVILVITVCIFONIFWSLITLIC
IHTVTLMSAGYFRIRSLADPDGPWKYPISYIAFHTYSIQGLLENEYVGTSFAGVGVRTISGYDALRNVYEVSPDINSKWKVNLVFLFMAVGLRIFVYILQLFRVAKNSSIRKFLRQNGSTN
NSR*

Fig. S1 The contig sequences of comp46786_c2_seq1 and comp78121_c0_seq10
Shown are the original contig sequences of comp46786_c2_seq1 and comp78121_c0_seq10 derived from transcriptome data. Deduced amino acid sequences were predicted using the ExPASy Translate tool (<https://web.expasy.org/translate/>). Reading frame 2 was used to predict the amino acid sequence of comp46786_c2_seq1. Red characters in the comp78121_c0_seq10 sequence represent the start codon “ATG” and the in-frame stop codons “TAG” and “TGA”; these are present prior to and following the ORF, respectively. The asterisk represents stop codon in the amino acid sequence.

(Fig. S2)



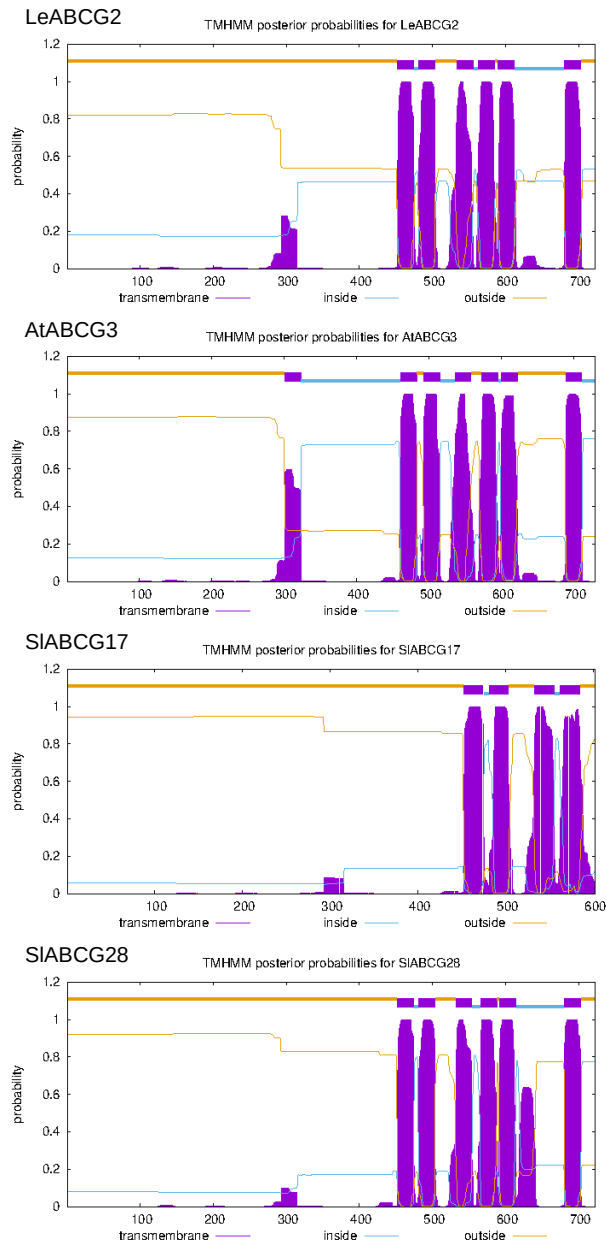


Fig. S2 Transmembrane helices of ABCG proteins as predicted by TMHMM

Shown are the results of TMHMM-2.0 prediction of the numbers and positions of transmembrane helices for LeABCG1, LeABCG2, and their orthologous proteins. Also shown are the predicted membrane topologies; purple boxes, blue lines, and orange lines indicate transmembrane helices and polypeptide topology inside and outside the membrane, respectively.

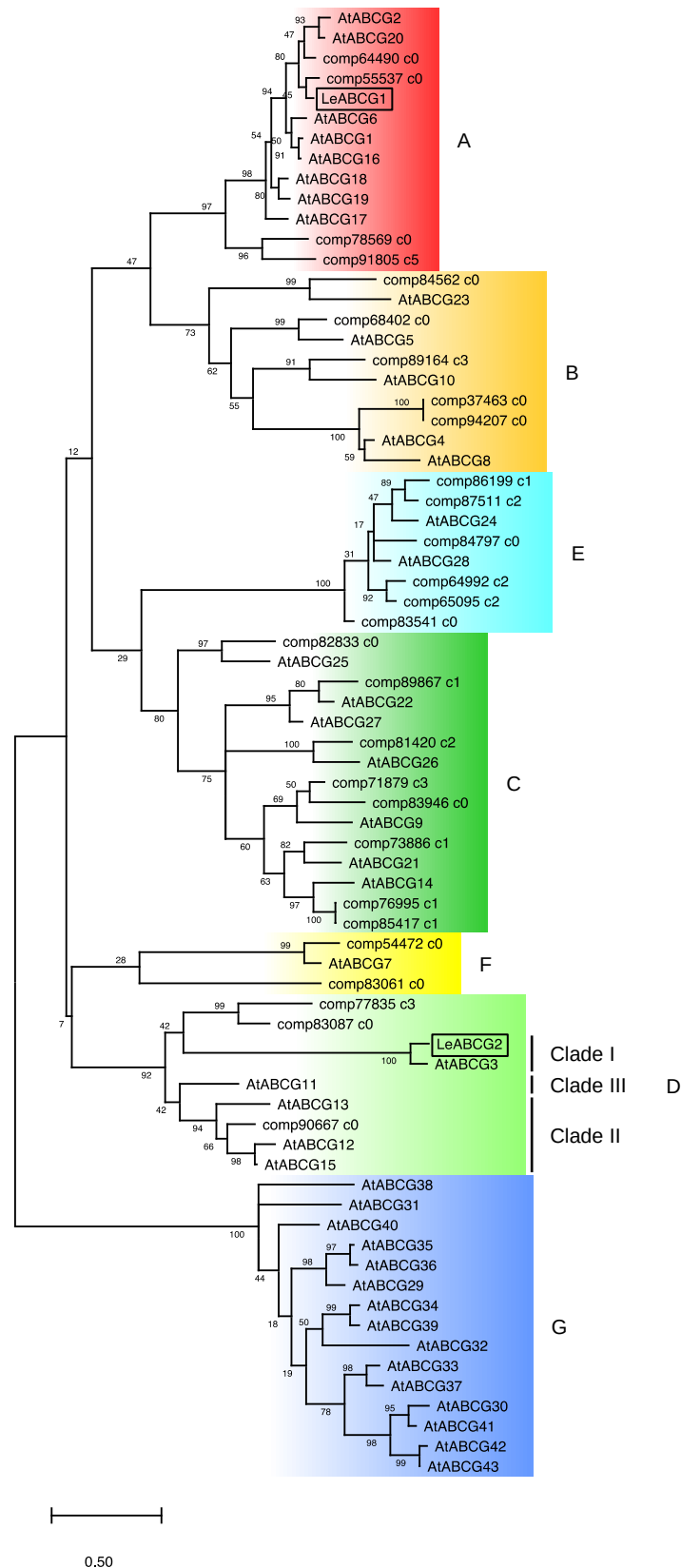


Fig. S3 Phylogenetic tree of *L. erythrorhizon* and *Arabidopsis* ABCG proteins

Polypeptide sequences of the NBD domain (i.e., between Walker A and Walker B) of ABCG proteins were used for alignment. The maximum likelihood method (i.e., the JTT model) was used to construct the tree. Bootstrapping used 1,000 replicates and the percentage of trees in which the associated taxa clustered together is shown next to the branches. Class (A–G) and clade (I–III) were classified as per Zhang et al. (2021) and Panikashvili et al. (2011).

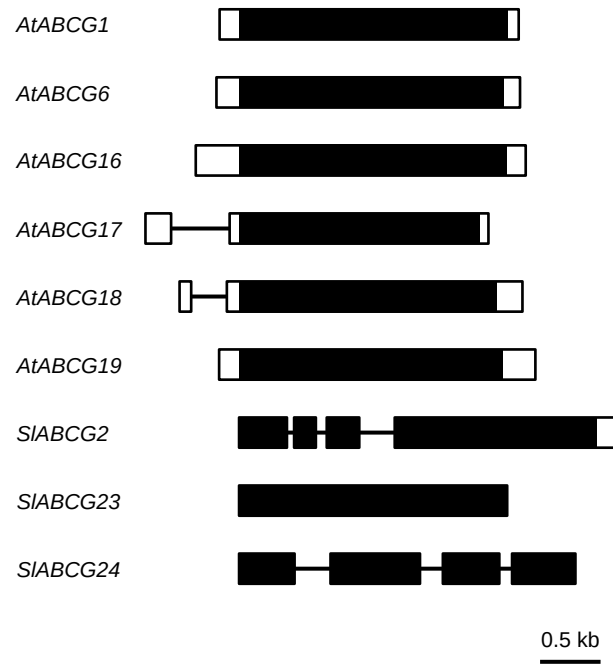


Fig. S4 Genomic structures of all Class A ABCG members from Arabidopsis and tomato
 Open boxes, closed boxes, and lines represent untranslated regions, exons, and introns, respectively.

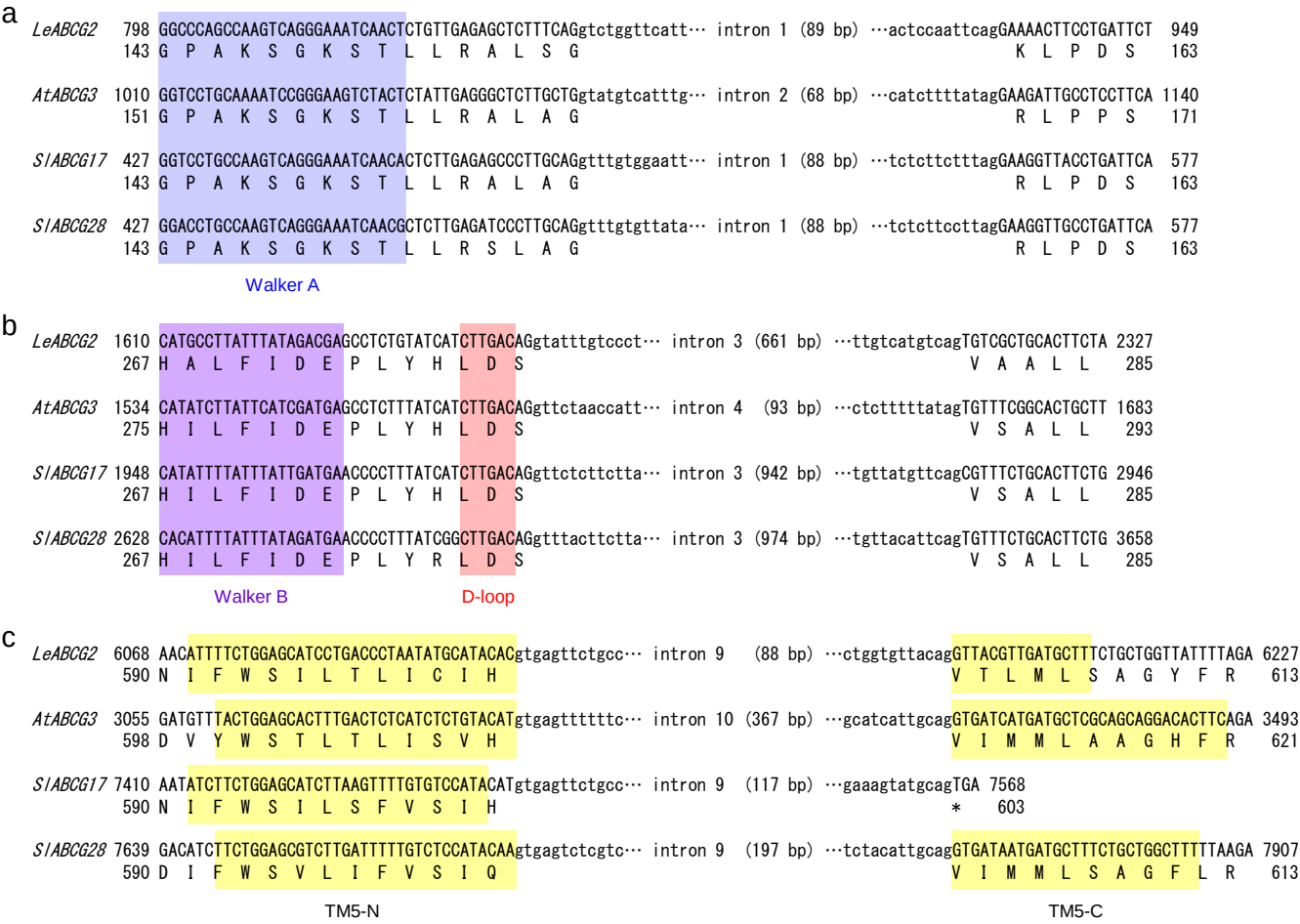


Fig. S5 The exon-intron junction sequence of *LeABCG2* and its orthologs

The positions of the first (a), third (b), and ninth introns (c) of the ORF region are shown in the alignment of the genomic sequences of *LeABCG2* and its Arabidopsis and tomato orthologs. Capital and lowercase letters indicate nucleotide sequences in exons and introns, respectively. Encoded amino acids are shown beneath the corresponding nucleotide sequence. The number of introns in the entire genomic sequence and the intron length are also shown. Conserved motifs are highlighted as follows: Walker A (blue), Walker B (purple), D-loop (red), N-terminal and C-terminal halves of the fifth transmembrane helix (TM5; yellow).

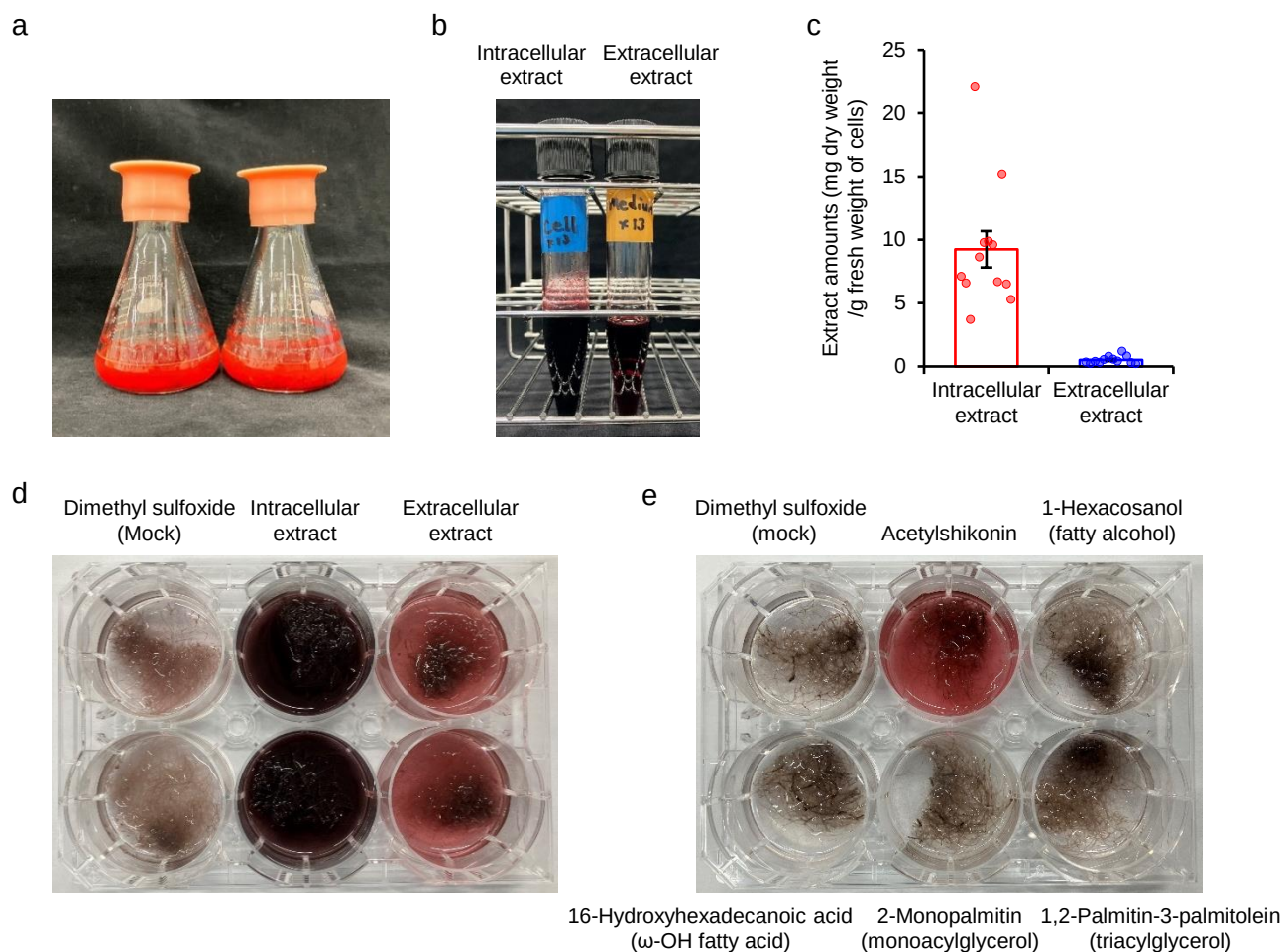
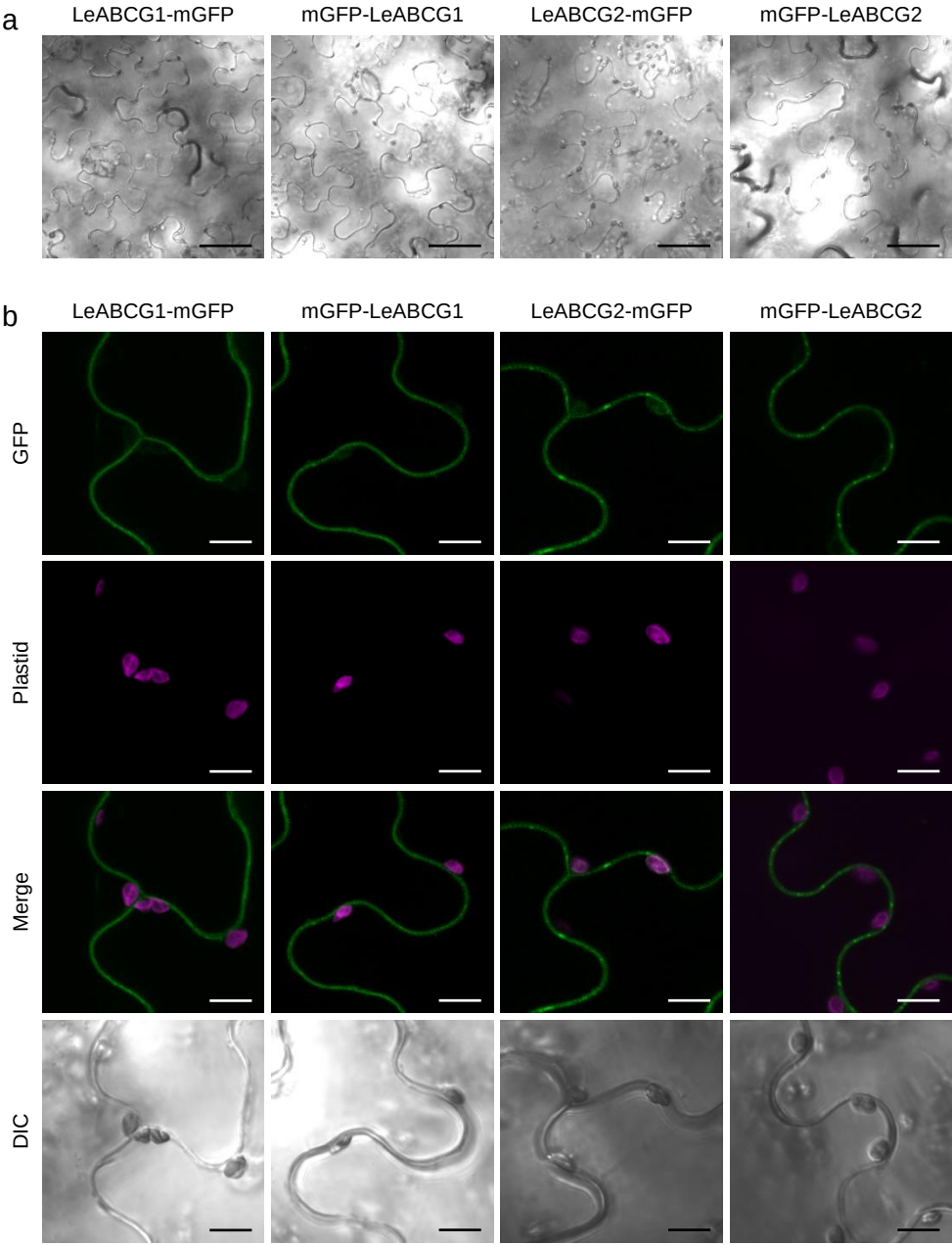


Fig. S6 Metabolite extracts of *L. erythrorhizon* and chemical treatment corresponding to Fig. 4

(a) *L. erythrorhizon* cells strain Mpr cultured in M9 medium for 20 days produce red pigment, shikonin derivatives. (b, c) Intracellular and extracellular metabolite solution (b) and the level of dried extracts (c) of Mpr cells extracted by chloroform-methanol. Bar graph indicates the mean with standard error of twelve independent cultures of the cells (intracellular and extracellular extracts are 9.25 ± 1.44 and 0.507 ± 0.084 mg dried extract per g fresh cell weight, respectively). (d) *L. erythrorhizon* hairy roots were treated with intracellular or extracellular metabolite extracts at 3 mg/mL or 0.3 mg/mL, respectively. (e) *L. erythrorhizon* hairy roots were treated with five extracellular metabolites produced by plants: 1 mM acetylshikonin, 1 mM C26:0 fatty alcohol (1-hexacosanol), 1 mM C16:0 ω -hydroxy fatty acid (16-hydroxyhexadecanoic acid), 1 mM monoacylglycerol (2-monopalmitin), and 0.1 mM triacylglycerol (1,2-palmitin-3-palmitolein). Dimethyl sulfoxide at 1.2% (v/v) in (d) and 2.0% (v/v) in (e) was used as mock solution.

(Fig. S7)



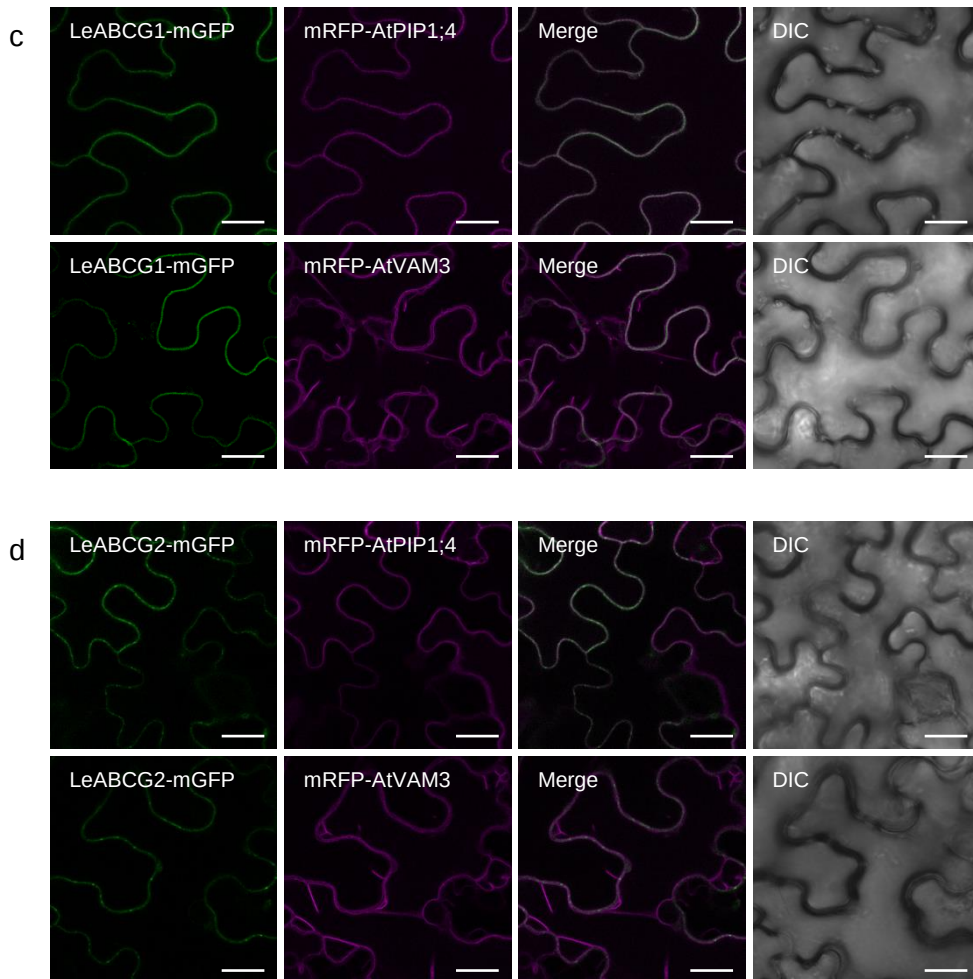


Fig. S7 Fluorescent micrographs of GFP fusion proteins for LeABCG1 and LeABCG2 corresponding to split images of Fig. 5

(a, b) Confocal images of *N. benthamiana* leaf epidermal cells transiently expressing GFP-fused LeABCG proteins. DIC images corresponding to the top panels of Fig. 5a are shown in **a**. Split and DIC images corresponding to the bottom panels of Fig. 5a are shown in **b**. Green, GFP fluorescence; magenta, auto-fluorescence of chlorophyll in plastids. (c, d) Confocal images of *N. benthamiana* leaf epidermal cells transiently coexpressing LeABCG1-mGFP (c) or LeABCG2-mGFP (d) with PM marker mRFP-AtPIP1;4 or tonoplast marker mRFP-AtVAM3. Split and DIC images corresponding to the left panels of Fig. 5b and Fig. 5c are shown in **c** and **d**, respectively. Green, GFP fluorescence; magenta, RFP fluorescence. Bars indicate 50 μ m (a), 10 μ m (b), and 20 μ m (c, d).

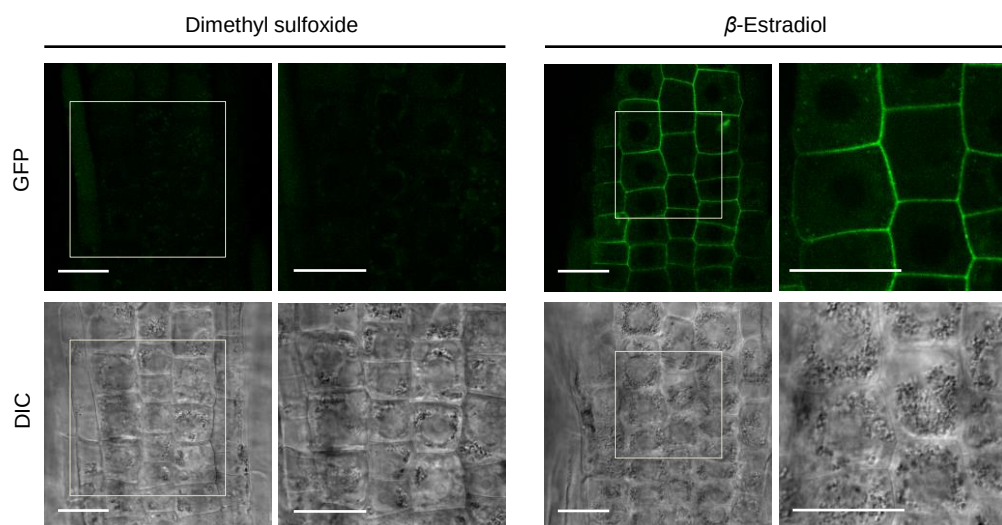


Fig. S8 Localization of LeABCG2-mGFP in *L. erythrorhizon* hairy roots

Shown is LeABCG2-mGFP expression in *L. erythrorhizon* hairy roots following transformation with pER8GW_LeABCG2-mGFP and after induction with 20 μ M β -estradiol for 20 hours. GFP fluorescence was detected as PM patterns in the hairy root cells in the presence of β -estradiol. Rectangles depict enlarged areas in right panels. Bars indicate 20 μ m. At least five hairy root lines were observed for each condition.

Table S1. Primers used in this study.

Purpose	Gene/vector	Sequence (5' to 3')
Sequence determination of cDNA/genomic	<i>LeABCG1</i>	ACATCTTTGTCCTATACTCCTGCCTCA
		AGAACTTGCATAAGCAGAATCATCAATAAAACC
Sequence determination of cDNA/genomic	<i>LeABCG2</i>	TTACTGGGGTTGTTAAAGATTTGAGC
		CATATCATCCATTCAAGTGAGGTTGG
Sequence determination of cDNA	<i>LeACT</i>	GCCAAAACCTTCAACTTAGTCAACCTCA
		CCATGGTACTAACCATACAAGAAACATTACC
Sequence determination of genomic	<i>LeACT</i>	ATGGCCGATGCCGAGGATATCCA
		TTAGAAGCACTTCCTGTGGACAATTGATGGA
Vector construction of pENTR_LeABCG1	<i>LeABCG1</i>	CACCATGTCTAATAACAATATGGTCCGAGAACATATCGTC
		TCAGTTCCTCTTGTTCTTACTCCCCAAC
		GTTCTCTTGTTCTTACTCCCCAACAAATAGAG
Vector construction of pENTR_LeABCG2 and pENTR_LeABCG2-mGFP	<i>LeABCG2</i>	GCAGGCTCCGCGGCCGCCACCATGGAAGAGATACAA TCACATTCTGATCATTATAGGTCTT
Vector construction of pENTR_LeABCG2	<i>LeABCG2</i>	AGCTGGGTCGGCGCGCCCTCATCTTGAATTGTTGT GCTTTGATTGCACC
		AGCTGGGTCGGCGCGCCCTCTTGAATTGTTGTGCT TTGATTGCACCG
		AGCTGGGTCGGCGCGCCCTTACTTGTACAGCTCGTC CATGCC
Vector construction of pENTR_AtPIP1;4	<i>AtPIP1;4</i>	GCAGGCTCCGCGGCCGCCACCATGGAAGGCAAAGA AGAAGATGTACGAGT
		AGCTGGGTCGGCGCGCCCTAACTCTTGCTCTTGAA AGGAATCGCTCT
		CACCATGAGTTTTCAAGATTTAGAATCAGGAAGAGGA TCAAGCTGCGAGTACTATAATCACGATCAG
Vector construction of pENTR_LeABCG2, pENTR_LeABCG2-mGFP, and pENTR_AtPIP1;4	<i>pENTR/D-TOPO</i>	CGCGCCGACCCAGCTTTCTTGACAAAGTT
		GGCCGCGGAGCCTGCTTTTTTGACAAAGT
Vector construction of pER8GW	Gateway cassette	GTCGACTCTAGCCTCGAACAAGTTTGTACAAAAAAGC TGAACG
		GGCCTGGATCGACTAGTAAGCGCTGTTATCAACCAC
RT-qPCR for <i>LeABCG1</i>	<i>LeABCG1</i>	TCACCCGCGAATATTCCTACTTAC
		CTTCTTTCAAAGTCGAACCTGGAG
RT-qPCR for <i>LeABCG2</i>	<i>LeABCG2</i>	CAGAGTTGCAGCAGTCTTCG
		CTCATCTCGCAGACCAACAA
RT-qPCR for <i>LeACT</i>	<i>LeACT</i>	GCCTTTATTGTCTGCATCACC
		GGGGAAAAATAGAAAGCCAAAAC

Table S2. Plasmid vectors used in this study.

Vector name	Type of vector	Purpose	References
pGEM_LeABCG1 cDNA	Subcloning	Determination of <i>LeABCG1</i> cDNA sequence	This study
pGEM_LeABCG1 genomic	Subcloning	Determination of <i>LeABCG1</i> genomic sequence	This study
pGEM_LeABCG2 cDNA	Subcloning	Determination of <i>LeABCG2</i> cDNA sequence	This study
pGEM_LeABCG2 genomic	Subcloning	Determination of <i>LeABCG2</i> genomic sequence	This study
pGEM_LeACT cDNA	Subcloning	Determination of <i>LeACT</i> cDNA sequence	This study
pGEM_LeACT genomic	Subcloning	Determination of <i>LeACT</i> genomic sequence	This study
WAVE 138Y		Cloning template of <i>AtPIP1;4</i>	Geldner et al. 2009
pENTR_LeABCG1-stop	Entry clone	Cloning of <i>LeABCG1</i> ORF	This study
pENTR_LeABCG1+stop	Entry clone	Cloning of <i>LeABCG1</i> ORF	This study
pENTR_LeABCG2-stop	Entry clone	Cloning of <i>LeABCG2</i> ORF	This study
pENTR_LeABCG2+stop	Entry clone	Cloning of <i>LeABCG2</i> ORF	This study
pENTR_AtPIP1;4	Entry clone	Cloning of <i>AtPIP1;4</i> ORF	This study
pENTR_AtVAM3	Entry clone	Cloning of <i>AtVAM3</i> ORF	This study
pGWB405m	Destination vector	-	Nakagawa et al. 2007; Segami et al. 2014
pGWB406m	Destination vector	-	Nakagawa et al. 2007; Segami et al. 2014
pGWB502	Destination vector	Cloning template of Gateway cassette	Nakagawa et al. 2007
pGWB555	Destination vector	-	Nakagawa et al. 2007
pER8	Destination vector	-	Zuo et al. 2000
pER8GW	Destination vector	-	This study
pGWB405m_LeABCG1	Expression vector	Localization analysis of LeABCG1-mGFP	This study
pGWB406m_LeABCG1	Expression vector	Localization analysis of mGFP-LeABCG1	This study
pGWB405m_LeABCG2	Expression vector	Localization analysis of LeABCG2-mGFP	This study
pGWB406m_LeABCG2	Expression vector	Localization analysis of mGFP-LeABCG2	This study
pER8GW_LeABCG2-mGFP	Expression vector	Localization analysis of LeABCG2-mGFP	This study
pGWB555_AtPIP1;4	Expression vector	Localization analysis of mRFP-AtAtPIP1;4	This study
pGWB555_AtVAM3	Expression vector	Localization analysis of mRFP-AtVAM3	This study

Table S3. Protein sequences of plant ABCG proteins used for phylogenetic analysis with corresponding accession numbers. Protein sequences of ABCG proteins from *Arabidopsis thaliana*, *Solanum lycopersicum*, and *Petunia x hybrida* were obtained from The Arabidopsis Information Resource (TAIR, <https://www.arabidopsis.org/>), Sol Genomics Network (SGN and ITAG4.0, <https://solgenomics.net/>), and UniProtKB (<https://www.uniprot.org/help/uniprotkb>), respectively.

Protein Name	Species	Database	Locus	Accession number
AtABCG1	<i>Arabidopsis thaliana</i>	TAIR	At2g39350	AT2G39350.1
AtABCG2	<i>Arabidopsis thaliana</i>	TAIR	At2g37360	AT2G37360.1
AtABCG3	<i>Arabidopsis thaliana</i>	TAIR	At2g28070	AT2G28070.1
AtABCG4	<i>Arabidopsis thaliana</i>	TAIR	At4g25750	AT4G25750.1
AtABCG5	<i>Arabidopsis thaliana</i>	TAIR	At2g13610	AT2G13610.1
AtABCG6	<i>Arabidopsis thaliana</i>	TAIR	At5g13580	AT5G13580.1
AtABCG7	<i>Arabidopsis thaliana</i>	TAIR	At2g01320	AT2G01320.3
AtABCG8	<i>Arabidopsis thaliana</i>	TAIR	At5g52860	AT5G52860.1
AtABCG9	<i>Arabidopsis thaliana</i>	TAIR	At4g27420	AT4G27420.1
AtABCG10	<i>Arabidopsis thaliana</i>	TAIR	At1g53270	AT1G53270.1
AtABCG11	<i>Arabidopsis thaliana</i>	TAIR	At1g17840	AT1G17840.1
AtABCG12	<i>Arabidopsis thaliana</i>	TAIR	At1g51500	AT1G51500.1
AtABCG13	<i>Arabidopsis thaliana</i>	TAIR	At1g51460	AT1G51460.1
AtABCG14	<i>Arabidopsis thaliana</i>	TAIR	At1g31770	AT1G31770.1
AtABCG15	<i>Arabidopsis thaliana</i>	TAIR	At3g21090	AT3G21090.1
AtABCG16	<i>Arabidopsis thaliana</i>	TAIR	At3g55090	AT3G55090.1
AtABCG17	<i>Arabidopsis thaliana</i>	TAIR	At3g55100	AT3G55100.1
AtABCG18	<i>Arabidopsis thaliana</i>	TAIR	At3g55110	AT3G55110.1
AtABCG19	<i>Arabidopsis thaliana</i>	TAIR	At3g55130	AT3G55130.1
AtABCG20	<i>Arabidopsis thaliana</i>	TAIR	At3g53510	AT3G53510.1
AtABCG21	<i>Arabidopsis thaliana</i>	TAIR	At3g25620	AT3G25620.2
AtABCG22	<i>Arabidopsis thaliana</i>	TAIR	At5g06530	AT5G06530.2
AtABCG23	<i>Arabidopsis thaliana</i>	TAIR	At5g19410	AT5G19410.1
AtABCG24	<i>Arabidopsis thaliana</i>	TAIR	At1g53390	AT1G53390.1
AtABCG25	<i>Arabidopsis thaliana</i>	TAIR	At1g71960	AT1G71960.1
AtABCG26	<i>Arabidopsis thaliana</i>	TAIR	At3g13220	AT3G13220.1
AtABCG27	<i>Arabidopsis thaliana</i>	TAIR	At3g52310	AT3G52310.1
AtABCG28	<i>Arabidopsis thaliana</i>	TAIR	At5g60740	AT5G60740.1
AtABCG29	<i>Arabidopsis thaliana</i>	TAIR	At3g16340	AT3G16340.1
AtABCG30	<i>Arabidopsis thaliana</i>	TAIR	At4g15230	AT4G15230.1
AtABCG31	<i>Arabidopsis thaliana</i>	TAIR	At2g29940	AT2G29940.1
AtABCG32	<i>Arabidopsis thaliana</i>	TAIR	At2g26910	AT2G26910.1
AtABCG33	<i>Arabidopsis thaliana</i>	TAIR	At2g37280	AT2G37280.1
AtABCG34	<i>Arabidopsis thaliana</i>	TAIR	At2g36380	AT2G36380.1
AtABCG35	<i>Arabidopsis thaliana</i>	TAIR	At1g15210	AT1G15210.1
AtABCG36	<i>Arabidopsis thaliana</i>	TAIR	At1g59870	AT1G59870.1
AtABCG37	<i>Arabidopsis thaliana</i>	TAIR	At3g53480	AT3G53480.1
AtABCG38	<i>Arabidopsis thaliana</i>	TAIR	At3g30842	AT3G30842.1
AtABCG39	<i>Arabidopsis thaliana</i>	TAIR	At1g66950	AT1G66950.1
AtABCG40	<i>Arabidopsis thaliana</i>	TAIR	At1g15520	AT1G15520.1
AtABCG41	<i>Arabidopsis thaliana</i>	TAIR	At4g15215	AT4G15215.1
AtABCG42	<i>Arabidopsis thaliana</i>	TAIR	At4g15233	AT4G15233.2
AtABCG43	<i>Arabidopsis thaliana</i>	TAIR	At4g15236	AT4G15236.1

(Table S3, continued)

Protein Name	Species	Database	Locus	Accession number
SIABCG1	<i>Solanum lycopersicum</i>	SGN	Solyc01g006720	Solyc01g006720.2.1
SIABCG2	<i>Solanum lycopersicum</i>	SGN	Solyc01g097430	Solyc01g097430.2.1
SIABCG3	<i>Solanum lycopersicum</i>	SGN	Solyc01g105450	Solyc01g105450.2.1
SIABCG4	<i>Solanum lycopersicum</i>	SGN	Solyc03g007690	Solyc03g007690.1.1
SIABCG5	<i>Solanum lycopersicum</i>	SGN	Solyc03g019760	Solyc03g019760.2.1
SIABCG6	<i>Solanum lycopersicum</i>	SGN	Solyc03g113690	Solyc03g113690.1.1
SIABCG7	<i>Solanum lycopersicum</i>	SGN	Solyc04g006960	Solyc04g006960.2.1
SIABCG8	<i>Solanum lycopersicum</i>	SGN	Solyc04g010200	Solyc04g010200.1.1
SIABCG9	<i>Solanum lycopersicum</i>	SGN	Solyc04g010210	Solyc04g010210.1.1
SIABCG10	<i>Solanum lycopersicum</i>	SGN	Solyc04g070970	Solyc04g070970.2.1
SIABCG11	<i>Solanum lycopersicum</i>	SGN	Solyc05g008350	Solyc05g008350.2.1
SIABCG12	<i>Solanum lycopersicum</i>	SGN	Solyc05g051530	Solyc05g051530.2.1
SIABCG13	<i>Solanum lycopersicum</i>	SGN	Solyc05g054890	Solyc05g054890.2.1
SIABCG14	<i>Solanum lycopersicum</i>	SGN	Solyc05g056470	Solyc05g056470.1.1
SIABCG15	<i>Solanum lycopersicum</i>	SGN	Solyc06g072090	Solyc06g072090.1.1
SIABCG16	<i>Solanum lycopersicum</i>	SGN	Solyc06g072100	Solyc06g072100.1.1
SIABCG17	<i>Solanum lycopersicum</i>	SGN	Solyc06g074970	Solyc06g074970.1.1
SIABCG18	<i>Solanum lycopersicum</i>	SGN	Solyc07g053300	Solyc07g053300.1.1
SIABCG19	<i>Solanum lycopersicum</i>	SGN	Solyc07g062630	Solyc07g062630.1.1
SIABCG20	<i>Solanum lycopersicum</i>	SGN	Solyc07g063400	Solyc07g063400.2.1
SIABCG21	<i>Solanum lycopersicum</i>	SGN	Solyc08g005580	Solyc08g005580.2.1
SIABCG22	<i>Solanum lycopersicum</i>	SGN	Solyc08g075430	Solyc08g075430.2.1
SIABCG23	<i>Solanum lycopersicum</i>	SGN	Solyc09g005970	Solyc09g005970.1.1
SIABCG24	<i>Solanum lycopersicum</i>	SGN	Solyc09g098410	Solyc09g098410.1.1
SIABCG25	<i>Solanum lycopersicum</i>	SGN	Solyc11g009100	Solyc11g009100.1.1
SIABCG26	<i>Solanum lycopersicum</i>	SGN	Solyc11g065350	Solyc11g065350.1.1
SIABCG27	<i>Solanum lycopersicum</i>	SGN	Solyc11g065360	Solyc11g065360.1.1
SIABCG28	<i>Solanum lycopersicum</i>	SGN	Solyc11g069710	Solyc11g069710.1.1
SIABCG29	<i>Solanum lycopersicum</i>	SGN	Solyc12g013630	Solyc12g013630.1.1
SIABCG30	<i>Solanum lycopersicum</i>	SGN	Solyc12g013640	Solyc12g013640.1.1
SIABCG31	<i>Solanum lycopersicum</i>	SGN	Solyc12g019620	Solyc12g019620.1.1
SIABCG32	<i>Solanum lycopersicum</i>	SGN	Solyc12g019640	Solyc12g019640.1.1
SIABCG33	<i>Solanum lycopersicum</i>	SGN	Solyc01g101070	Solyc01g101070.2.1
SIABCG34	<i>Solanum lycopersicum</i>	SGN	Solyc02g081870	Solyc02g081870.2.1
SIABCG35	<i>Solanum lycopersicum</i>	SGN	Solyc03g120980	Solyc03g120980.2.1
SIABCG36	<i>Solanum lycopersicum</i>	SGN	Solyc05g018510	Solyc05g018510.2.1
SIABCG37	<i>Solanum lycopersicum</i>	SGN	Solyc05g053570	Solyc05g053570.2.1
SIABCG38	<i>Solanum lycopersicum</i>	SGN	Solyc05g053590	Solyc05g053590.2.1
SIABCG39	<i>Solanum lycopersicum</i>	SGN	Solyc05g053600	Solyc05g053600.2.1
SIABCG40	<i>Solanum lycopersicum</i>	SGN	Solyc05g053610	Solyc05g053610.2.1

(Table S3, continued)

Protein Name	Species	Database	Locus	Accession number
SIABCG41	<i>Solanum lycopersicum</i>	SGN	Solyc05g055330	Solyc05g055330.2.1
SIABCG42	<i>Solanum lycopersicum</i>	SGN	Solyc06g065670	Solyc06g065670.2.1
SIABCG43	<i>Solanum lycopersicum</i>	SGN	Solyc06g076930	Solyc06g076930.1.1
SIABCG44	<i>Solanum lycopersicum</i>	SGN	Solyc08g067610	Solyc08g067610.2.1
SIABCG45	<i>Solanum lycopersicum</i>	SGN	Solyc08g067620	Solyc08g067620.2.1
SIABCG46	<i>Solanum lycopersicum</i>	SGN	Solyc09g091660	Solyc09g091660.2.1
SIABCG47	<i>Solanum lycopersicum</i>	SGN	Solyc09g091670	Solyc09g091670.2.1
SIABCG48	<i>Solanum lycopersicum</i>	SGN	Solyc11g007280	Solyc11g007280.1.1
SIABCG49	<i>Solanum lycopersicum</i>	SGN	Solyc11g007290	Solyc11g007290.1.1
SIABCG50	<i>Solanum lycopersicum</i>	SGN	Solyc11g007300	Solyc11g007300.1.1
SIABCG51	<i>Solanum lycopersicum</i>	SGN	Solyc11g067000	Solyc11g067000.1.1
SIABCG52	<i>Solanum lycopersicum</i>	SGN	Solyc12g098210	Solyc12g098210.1.1
SIABCG53	<i>Solanum lycopersicum</i>	SGN	Solyc12g100180	Solyc12g100180.1.1
SIABCG54	<i>Solanum lycopersicum</i>	SGN	Solyc12g100190	Solyc12g100190.1.1
SIABCG55	<i>Solanum lycopersicum</i>	SGN	Solyc00g233480	Solyc00g233480.1.1
SIABCG56	<i>Solanum lycopersicum</i>	SGN	Solyc01g105400	Solyc01g105400.2.1
SIABCG57	<i>Solanum lycopersicum</i>	SGN	Solyc04g025170	Solyc04g025170.2.1
SIABCG58	<i>Solanum lycopersicum</i>	SGN	Solyc05g051540	Solyc05g051540.1.1
SIABCG59	<i>Solanum lycopersicum</i>	SGN	Solyc06g036240	Solyc06g036240.1.1
SIABCG60	<i>Solanum lycopersicum</i>	SGN	Solyc06g075020	Solyc06g075020.2.1
SIABCG61	<i>Solanum lycopersicum</i>	SGN	Solyc07g065770	Solyc07g065770.2.1
SIABCG62	<i>Solanum lycopersicum</i>	SGN	Solyc09g008000	Solyc09g008000.2.1
SIABCG63	<i>Solanum lycopersicum</i>	SGN	Solyc11g018690	Solyc11g018690.1.1
SIABCG64	<i>Solanum lycopersicum</i>	SGN	Solyc11g069820	Solyc11g069820.1.1
SIABCG65	<i>Solanum lycopersicum</i>	SGN	Solyc07g065780	Solyc07g065780.1.1
SIABCG66	<i>Solanum lycopersicum</i>	SGN	Solyc11g018680	Solyc11g018680.1.1
SIABCG67	<i>Solanum lycopersicum</i>	SGN	Solyc00g164680	Solyc00g164680.1.1
SIABCG68	<i>Solanum lycopersicum</i>	SGN	Solyc02g055530	Solyc02g055530.2.1
SIABCG69	<i>Solanum lycopersicum</i>	SGN	Solyc04g076170	Solyc04g076170.1.1
SIABCG70	<i>Solanum lycopersicum</i>	SGN	Solyc09g042280	Solyc09g042280.1.1
PhABCG1	<i>Petunia x hybrida</i>	UniProtKB /Swiss-Prot	-	H9BZ66.1

Table S4. FPKM values of selected contigs annotated as membrane transporters.

Contigs annotated as membrane transporters are selected if, according to gene expression analyses, their FPKM values showed more than threefold increase in shikonin-producing tissues (i.e., root barks, hairy roots, and cells cultured in the dark). In total, 23 contigs (including 20 non-redundant genes) are annotated as membrane transporters. Annotations of the BLAST top hit proteins in The Arabidopsis Information Resource (TAIR) and NCBI databases is classified in the “Protein family” column.

Contig ID	Protein family	Intact tissue		Hairy roots		Cultured cells	
		Peeled roots	Root barks	Light	Dark	Light	Dark
comp114549_c0_seq1	Choline transporter	0	0.90	0	3.27	0	1.04
comp18047_c0_seq2	Aquaporin	0	2.68	0	3.13	0	1.66
comp3484_c1_seq1	Amino acid transporter	0	0.27	0	0.11	0.10	0.33
comp46786_c2_seq1	ABCG/WBC transporter	0	11.76	0	1.29	0	1.23
comp53960_c0_seq1	Magnesium transporter	0.08	0.29	0	0.16	0	0.26
comp56811_c0_seq1	Dicarboxylate/tricarboxylate carrier	0	0.09	0	0.11	0	0.22
comp58108_c0_seq1	Choline transporter	0	0.14	0	0.08	0	0.43
comp60015_c0_seq1	Dicarboxylate carrier	0.29	1.40	0	0.30	0	0.15
comp62710_c1_seq1	HCO ₃ ⁻ transporter	0	0.20	0	0.47	0	0.46
comp73638_c0_seq1	NRT1/PTR family	0.22	11.62	19.96	68.10	11.45	104.76
comp73638_c0_seq2	NRT1/PTR family	0.22	11.62	19.96	68.10	11.45	104.76
comp73638_c6_seq1	NRT1/PTR family	0.16	8.60	15.29	60.37	7.33	131.96
comp73638_c6_seq2	NRT1/PTR family	0.16	8.60	15.29	60.37	7.33	131.96
comp73638_c6_seq3	NRT1/PTR family	0.16	8.60	15.29	60.37	7.33	131.96
comp73998_c1_seq2	ABCB/MDR/PGP transporter	0	0.38	0.10	1.22	0	0.03
comp74687_c2_seq1	Sulfate transporter	0	2.14	0	5.14	0	4.92
comp77047_c2_seq1	Na ⁺ /H ⁺ antiporter	0	2.67	0	1.59	0.81	3.66
comp81225_c1_seq1	NRT1/PTR family	0.02	0.07	0.04	0.15	0.08	0.29
comp85375_c9_seq1	Phosphate transporter	0	1.14	0	0.69	0	0.65
comp87715_c4_seq1	Major facilitator superfamily	0	0.93	0	0.56	0	0.53
comp89353_c2_seq1	Na ⁺ /Ca ²⁺ exchanger	0	3.44	0	6.11	0	10.00
comp89522_c15_seq1	Amino acid transporter	0	0.70	0	0.85	0	0.40
comp92896_c0_seq1	NRT1/PTR family	0.08	9.23	0.19	14.05	0.08	48.22

Table S5. BLAST hits for *L. erythrorhizon* genes homologous to the PhABCG1 protein. The results of a homology search using TBLASTN with the PhABCG1 protein sequence as a query and *L. erythrorhizon* transcriptome data as a reference. Twenty-one contigs corresponding to six genes were selected using a cut off set to e-70.

Contig ID	Contig length	Score (bits)	E Value	
comp90667_c0_seq2	2,535	610	e-174	
comp77835_c0_seq3	1,521	564	e-160	
comp83087_c0_seq1	1,202	382	e-121	
comp78121_c0_seq10	3,491	424	e-118	<i>LeABCG2</i>
comp78121_c0_seq6	2,973	424	e-118	
comp78121_c0_seq3	3,014	424	e-118	
comp78121_c0_seq2	2,988	424	e-118	
comp78121_c0_seq1	2,947	424	e-118	
comp90667_c0_seq1	2,095	389	e-117	
comp51814_c0_seq1	1,467	416	e-115	
comp83087_c0_seq4	1,080	388	e-107	
comp83087_c0_seq3	993	333	1e-90	
comp83061_c0_seq1	2,225	323	1e-87	
comp77835_c0_seq1	792	305	6e-82	
comp83061_c0_seq3	2,313	303	1e-81	
comp78121_c0_seq9	3,490	285	5e-76	
comp78121_c0_seq8	3,423	285	5e-76	
comp78121_c0_seq7	3,967	285	5e-76	
comp78121_c0_seq5	3,464	285	5e-76	
comp78121_c0_seq4	3,449	285	5e-76	
comp77835_c0_seq4	625	160	8e-72	

Table S6. Gene expression of contigs encoding half-sized ABCG transporters in *L. erythrorhizon*. Expression levels (FPKM values) of genes encoding a half-sized ABCG transporter are shown. Six non-redundant genes were selected via homology search of PhABCG1; these are listed in the column named as “Table S5”. Non-redundant genes encoding half-sized ABCG transporters that possess a NBD domain (i.e., Walker A to Walker B) are shown in the column named as “Fig. 1a”. The gene comp46786_c2 is identical to contig comp46786_c2_seq1, which corresponds to *LeABCG1*. The gene comp78121_c0 includes the contig comp78121_c0_seq10, which corresponds to *LeABCG2*.

Contig ID	Intact tissue		Hairy roots		Cultured cells		Table S5	Fig. 1a
	Peeled roots	Root bark	Light	Dark	Light	Dark		
comp37463_c0	0.99	0.28	0.18	0.51	0.04	0.12		○
comp46786_c2	0	11.76	0	1.29	0	1.23		○ (<i>LeABCG1</i>)
comp51814_c0	0.25	0.21	0.03	0.11	1.19	1.85	○	
comp54472_c0	380.08	367.30	28.39	49.84	14.76	12.69		○
comp55537_c0	0.15	54.34	16.99	21.19	1.31	1.44		○
comp64490_c0	0.27	6.53	10.40	13.50	0.38	0.24		○
comp64992_c2	0	0	0	0.02	0.34	0.53		○
comp65095_c2	0.78	0.57	0.03	0.03	0.01	0.01		○
comp68402_c0	1.09	0.68	0.36	0.20	0.03	0.03		○
comp71879_c3	0.51	4.72	0.32	0.39	0.34	0		○
comp73886_c1	18.03	6.13	9.09	17.11	6.57	1.82		○
comp76995_c1	0.20	0.04	0.23	0.29	0.70	0.24		○
comp77835_c0	1.96	3.32	0.01	0.13	0.87	0.03	○	
comp77835_c3	10.49	10.10	0.27	0.81	1.20	0.14		○
comp78121_c0	32.47	17.88	22.18	24.77	81.11	83.78	○	○ (<i>LeABCG2</i>)
comp78569_c0	0.15	0.05	0.11	0.05	0.25	0.20		○
comp81420_c2	1.40	0.99	32.04	18.07	3.78	3.58		○
comp82833_c0	1.25	0.59	0.38	0.65	0.21	0.04		○
comp83061_c0	0.03	6.02	0	0	0	0	○	○
comp83087_c0	0.30	0.11	0.39	0.47	0.09	0.13	○	○
comp83541_c0	0.66	1.78	4.96	2.87	0.24	0.46		○
comp83946_c0	0.10	0.04	1.07	1.18	1.10	1.78		○
comp84562_c0	0.04	2.09	2.89	4.90	0	0.04		○
comp84797_c0	1.94	0.48	0.44	4.39	0	0.01		○
comp85417_c1	11.36	2.39	34.07	58.91	4.17	2.79		○
comp86199_c1	1.28	0.02	0.64	0.15	4.32	3.72		○
comp87511_c2	4.22	3.73	8.07	6.08	7.03	8.93		○
comp89164_c3	74.99	79.25	70.40	91.39	95.52	77.83		○
comp89867_c1	2.18	1.91	2.28	2.55	1.16	0.99		○
comp90667_c0	2.18	1.55	1.12	1.07	0.77	0.73	○	○
comp91805_c5	0.11	0	0.20	0.24	0.62	0.36		○
comp94207_c0	1.33	0.43	3.72	2.02	0.78	1.05		○

Table S7. Sequence identity of LeABCG1 or LeABCG2 and their orthologs. Amino acid sequence identities among whole polypeptides of LeABCG1 orthologs (**a**), NBD domains of LeABCG1 orthologs (**b**), whole polypeptides of LeABCG2 orthologs (**c**), and NBD domains of LeABCG2 orthologs (**d**) were calculated using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). The percentage identity between two sequences is shown in each column.

a		LeABCG1	AtABCG2	AtABCG20	SIABCG8	SIABCG9	SIABCG13
	LeABCG1	-					
	AtABCG2	66.30	-				
	AtABCG20	69.11	82.72	-			
	SIABCG8	75.10	68.30	70.07	-		
	SIABCG9	73.56	66.67	68.52	87.82	-	
	SIABCG13	76.64	68.29	70.76	78.32	76.79	-

b		LeABCG1	AtABCG2	AtABCG20	SIABCG8	SIABCG9	SIABCG13
	LeABCG1	-					
	AtABCG2	87.77	-				
	AtABCG20	89.21	94.24	-			
	SIABCG8	92.09	87.77	88.49	-		
	SIABCG9	90.65	87.05	87.77	97.12	-	
	SIABCG13	89.93	87.77	88.49	92.09	92.09	-

c		LeABCG2	AtABCG3	SIABCG17	SIABCG28
	LeABCG2	-			
	AtABCG3	76.73	-		
	SIABCG17	80.40	79.90	-	
	SIABCG28	82.02	79.22	87.54	-

d		LeABCG2	AtABCG3	SIABCG17	SIABCG28
	LeABCG2	-			
	AtABCG3	89.05	-		
	SIABCG17	87.59	87.59	-	
	SIABCG28	85.40	89.78	90.51	-