Journal of Plant Research
https://doi.org/10.1007/s10265-021-01257-9

REGULAR PAPER – PHYSIOLOGY/BIOCHEMISTRY/MOLECULAR AND CELLULAR BIOLOGY

ADAP is a possible negative regulator of glucosinolate biosynthesis in *Arabidopsis thaliana* based on clustering and gene expression analyses

S. Harun[1] · E. R. Rohani[2] · M. Ohme-Takagi[3,4] · H.-H. Goh[2] · Z.-A. Mohamed-Hussein[1,5]

Received: 4 August 2020 / Accepted: 25 January 2021

Abstract
Glucosinolates (GSLs) are plant secondary metabolites consisting of sulfur and nitrogen, commonly found in *Brassicaceae* crops, such as *Arabidopsis thaliana*. These compounds are known for their roles in plant defense mechanisms against pests and pathogens. 'Guilt-by-association' (GBA) approach predicts genes encoding proteins with similar function tend to share gene expression pattern generated from high throughput sequencing data. Recent studies have successfully identified GSL genes using GBA approach, followed by targeted verification of gene expression and metabolite data. Therefore, a GSL co-expression network was constructed using known GSL genes obtained from our in-house database, SuCComBase. DPClusO was used to identify subnetworks of the GSL co-expression network followed by Fisher's exact test leading to the discovery of a potential gene that encodes the ARIA-interacting double AP2-domain protein (ADAP) transcription factor (TF). Further functional analysis was performed using an effective gene silencing system known as CRES-T. By applying CRES-T, ADAP TF gene was fused to a plant-specific EAR-motif repressor domain (SRDX), which suppresses the expression of ADAP target genes. In this study, ADAP was proposed as a negative regulator in aliphatic GSL biosynthesis due to the over-expression of downstream aliphatic GSL genes (*UGT74C1* and *IPMI1*) in ADAP-SRDX line. The significant over-expression of *ADAP* gene in the ADAP-SRDX line also suggests the behavior of the TF that negatively affects the expression of *UGT74C1* and *IPMI1* via a feedback mechanism in *A. thaliana*.

Keywords Clustering · CRES-T · Gene network · Glucosinolates · Negative regulation · Secondary metabolites · SRDX

Introduction

Glucosinolates (GSLs) play an important role in plant defense, especially in the *Brassicaceae* crops which includes *Arabidopsis thaliana* (L.) Heynh. (Arabidopsis). These sulfur- and nitrogen-containing compounds and their degradation products play essential roles in the resistance of plants against herbivores and pathogens (Clay et al. 2009). All GSLs have a similar general chemical structure and are categorized into three types according to side chains (R), which corresponds to different amino acid derivatives. They are aliphatic GSLs from methionine (Met), indolic GSLs from tryptophan (Trp), and benzyl GSLs from phenylalanine (Phe) and tyrosine (Tyr). GSLs are responsible for the bitter flavors of *Brassicaceae* vegetables, including turnip (*B. rapa* ssp. *rapa*), broccoli (*B. oleracea* var. *italica*), and cauliflower (*B. oleracea* var. *botrytis*) (Padilla et al. 2007).

✉ Z.-A. Mohamed-Hussein
zeti.hussein@ukm.edu.my

1 Centre for Bioinformatics Research, Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

2 Centre for Plant Biotechnology, Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

3 Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, Japan

4 Graduate School of Science and Engineering, Saitama University, Saitama, Japan

5 Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Published online: 08 February 2021

Several types of GSLs have been studied for their cancer-preventing properties. Glucotropaeolin (benzyl GSL) and glucobrassicin (indolic GSL) produce benzyl isothiocyanate (BITC) and indole-3-carbinol, respectively. These products were found to suppress the growth of mammalian carcinogenic cells (Choi et al. 2010; Rao 2013). A previous study conducted in *B. oleracea* found a significant increment of glucotropaeolin concentrations in organic plants compared to the conventional breeding approach (Rossetto et al. 2013). Glucotropaeolin content was undetectable in the wild-type Arabidopsis plants and only reported in the rosette leaves of transgenic lines expressing CYP79A2 with CaMV35S promoter (Wittstock and Halkier 2000). Meanwhile, glucobrassicin was found in all *Brassicaceae* plants such as *B. oleracea* (Choi et al. 2014), *B. rapa* (Chung et al. 2016; Padilla et al. 2007), and Arabidopsis specifically in the roots and mature rosette leaves (Brown et al. 2003).

In general, GSL biosynthesis comprises several groups of genes initiated by the transcription factors (TFs) which regulate the production of various secondary metabolites from GSL derivatives. In Arabidopsis, GSL regulation involved six different MYBs originated from the R2R3-MYB TF family. MYB34, MYB51, and MYB122 control the production of indolic GSLs, whereas MYB28, MYB29, and MYB76 control the production of aliphatic GSLs (Frerigmann and Gigolashvili 2014; Gigolashvili et al. 2008). TF involved in the regulation of benzyl GSLs is still unknown. The main step in GSL biosynthesis is the core structure synthesis, in which most of the biosynthetic genes involved in indolic and benzyl GSLs are similar. In indolic GSL biosynthesis, CYP79B3 (tryptophan *N*-monooxygenase 2) catalyzes tryptophan derivatives whereas CYP79A2 (phenylalanine *N*-monooxygenase) catalyzes the phenylalanine substrate in the benzyl GSL biosynthesis (Wittstock and Halkier 2000).

CYP83B1 (CYP83B1 monooxygenase) then converts both tryptophan- and phenylalanine-derived acetaldoximes into aci-nitro compounds (Naur et al. 2003). The remaining steps of GSL core structure formation would involve several GSL biosynthetic genes that accommodate all GSL precursors regardless of their associated side chains. In this step, the S-alkylthiohydroximates are converted to thiohydroximic acids in a reaction catalyzed by SUR1 (C–S lyase) (Mikkelsen et al. 2004). In the glycosylation process, UGT74B1 was suggested to metabolize thiohydroximates based on the in vitro and in vivo analysis performed on this enzyme (Grubb et al. 2004). The final step in GSL core structure synthesis is the sulfation process of the desulfoglucosinolates to form intact GSLs involving the cytosolic sulfotransferase group such as SOT16, SOT17, and SOT18. Biochemical characterization of sulfotransferases suggests the role of SOT16 to metabolize phenylalanine- and tryptophan-derived desulfoglucosinolates, such as glucobrassicin (3-indolylmethyl GSL) (Klein and Papenbrock 2009). The continuous effort of collecting all molecules related to GSL biosynthesis led to the development of SuCCombase (Harun et al. 2019) that was used in this study. A recent review article reported a total of 113 GSL genes that can be categorized into TFs, biosynthetic genes, and protein transporters that have experimental evidences in the last 20 years (Harun et al. 2020).

'Guilt-by-association' (GBA) principle was previously used to identify novel GSL-related genes encoding TFs (Gigolashvili et al. 2007; Hirai et al. 2007; Sønderby et al. 2007) and enzymes (Geu-Flores et al. 2009; Knill et al. 2008; Sawada et al. 2009). This principle proposed that genes that have similar functions tend to be connected and can be used to predict new functions for the respective gene encoding proteins (Liesecke et al. 2018). The GBA principle is also known as fundamental in constructing and analyzing gene co-expression networks in functional genomics studies (Gillis and Pavlidis 2012; Wong 2020). Thus, by applying this principle, the co-expressed genes that contain information of known gene function will be used as a reference to determine the function of the genes of interest.

In this study, the focus is to discover TF related to the GSL pathway to elucidate the regulation of GSL biosynthesis. Thus, a GSL co-expression was constructed, and a graph clustering approach identified a potential gene encoding TF, ARIA-interacting double AP2-domain protein (ADAP) from the calculated significant subnetworks. ADAP belongs to the APETALA2/Ethylene Responsive Factor (AP2/ERF) TF family that consists of 147 TFs according to their DNA-binding domain sequences called AP2/ERF domain (Broekgaarden et al. 2015; Huang et al. 2015). ADAP is also known as ARIA-interacting double AP2-domain protein, which is found to regulate seed germination, seedling growth, ABA responses, and several abiotic factors such as salt and drought through in vivo functional analysis (Lee et al. 2009).

Detailed gene expression pattern and sequence analyses were conducted on ADAP and target GSL genes to investigate the regulatory mechanism of ADAP in GSL biosynthesis. The overall expression of *ADAP* and known GSL genes, which include TFs (*MYB28, MYB29*), and biosynthetic genes (*IMDH1, IPMI2 MAM1, BCAT4*, *CYP79F2, CYP83A1, GSTU20, GSTF11, UGT74C1, FMOGS-OX3*, and *AOP2*) showed high expression in chemical stress conditions. *Cis*-regulatory sequence analysis was performed to understand the regulatory mechanism of ADAP and UGT74C1, where they share similar TF binding sites (TFbs) with MYC2, MYC4, and OBP2 TFs.

Based on the preliminary bioinformatics analysis of ADAP's contribution in regulating GSL biosynthesis, chimeric repressor gene silencing technology (CRES-T) was applied on ADAP TF for functional analysis. CRES-T is a useful tool for functional analysis, specifically in plant TFs (Hiratsu et al. 2003; Mitsuda et al. 2011). This approach utilizes a plant-specific EAR-motif repressor domain

(SRDX)-motif to the C-terminal of the TF into dominant negative regulators that repress downstream target genes (Cen et al. 2016; Kazama et al. 2013). In this study, ADAP-SRDX line was obtained and an over-expression of downstream aliphatic GSL genes (*UGT74C1* and *IPMI1*) was found in the transgenic Arabidopsis plants using qPCR. Hence, ADAP was suggested as a negative regulator in aliphatic GSL biosynthesis via a feedback mechanism due to the significant over-expression of ADAP gene in Arabidopsis ADAP-SRDX line.

Materials and methods

Data collection and gene-co-expression network construction

A comprehensive pathway database search was carried out using SuCComBase (http://plant-scc.org/) (Harun et al. 2019), which is a manually curated sulfur-containing compound database. SuCComBase compiles known glucosinolate (GSL) genes from literature and pathway databases such as the Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/) (Kanehisa et al. 2016) and AraCyc (https://www.arabidopsis.org/biocyc/) (Mueller et al. 2003). Several relevant keywords were used, such as "glucosinolate" and "GSL". The known GSL genes were used as a query against four co-expression tools, which are ATTED (Aoki et al. 2016), AraNet v2 (Lee et al. 2014), GeneMANIA (Montojo et al. 2014; Warde-Farley et al. 2010) and STRING (Szklarczyk et al. 2019) to identify "additional" co-expressed genes. The "additional" genes are hypothesized as potential GSL genes based on the 'guilt-by-association' principle. An integrated gene network was produced using Cytoscape 3.7.1 (Shannon et al. 2003).

Calculating sub-networks and Fisher's exact test

The gene co-expression network was analyzed using DPClusOST (Karim et al. 2017), which utilizes DPClusO algorithm (Altaf-Ul-Amin et al. 2006, 2012), in which 0.6 density value and OV Coff 0.15 were selected. The clustering algorithm is useful for an undirected graph consisting of a finite set of nodes N and a finite set of edges E. In this algorithm, density d_k and cluster property cp_{nk} are two essential parameters that are introduced. Density d_k of cluster k refers to the ratio of the number of actual cluster edges ($|E|$) and the maximum possible number of cluster edges ($|E|_{max}$). Thus, cluster property cp_{nk} is represented by the following equation:

$$\mathrm{cp_{nk}} = \frac{E_{nk}}{d_k \times N_k}$$

N_k refers to the number of nodes in cluster k. E_{nk} is the total number of edges between the nodes n in the cluster.

Each cluster generated was then evaluated using Fisher's exact test (Fisher 1992) to assess GSL gene enrichment in each of the identified clusters. Table 1 shows the values of *a, b, c* and *d* introduced in this study.

Expression pattern analysis and sequence analysis of potential GSL genes

ePlant (https://bar.utoronto.ca/eplant/) was used to visualize the expression levels across various conditions collected from more than 350 samples in Arabidopsis (Waese et al. 2017). Expression Angler (http://bar.utoronto.ca/ExpressionAngler/) was used to identify the expression of potential genes and known GSL genes under several chemical stresses, such as jasmonic acid (JA), salicylic acid (SA), abscisic acid (ABA), and ethylene (ET). Next, PlantPAN 3.0 (http://plantpan2.itps.ncku.edu.tw/) was used to identify specific *cis*-regulatory sequences in the promoter regions of the potential and other known GSL genes. This is to further validate the potential of TF (ADAP) in regulating the GSL biosynthetic pathway.

Plant materials and growth conditions

The wild-type (WT) Arabidopsis Columbia-0 (Col-0) seeds were used as a control. Seeds were surface sterilized for ten minutes using 20% commercial bleach with addition of 0.05% tween 20, before rinsing three times with sterile distilled water. For stratification, seeds were imbibed in distilled water for three days in the dark at 4 °C to brake dormancy. Next, T_0 seeds were grown in a growth chamber (Percival Scientific, USA) with a 16 h light/ 8 h dark photoperiod. Day/ night temperatures were at 22 °C / 20 °C, and relative humidity was 60%. Plants were grown on ½ Murashige and Skoog (MS) medium (Murashige and Skoog 1962). Transgenic Arabidopsis plants that express the *35S:ADAP-SRDX* construct was prepared by fusion of the plant-specific repression domain SRDX (SUPERMAN REPRESSION DOMAIN X; LDLDLELRLGFA) to the 3′ end of the coding region of ADAP.

Table 1 The contingency table prepared in this study to calculate known GSL gene enrichment clusters

	Glucosinolate genes	Non-glucosinolate genes	
In cluster	*a*	*b*	*a* + *b*
Not in cluster	*c*	*d*	*c* + *d*
	a + *c*	*b* + *d*	*n*

[a]Here n refer to the total number of genes in the gene network

Primer design and seed selection

All primer sets used for PCR and reverse-transcription quantitative real-time PCR (RT-qPCR) are listed in Table S1. PrimerBlast software was used to design the following primers in producing first strand products and synthesized commercially (First BASE Laboratories). The integration of 35S:ADAP-SRDX transgenes were confirmed through PCR. Hygromycin was used in selecting the transformed ADAP-SRDX plants. Positive lines were grown to T_2 generation to obtain homozygous seeds used in this study.

Morphological analysis

The number of leaf and rosette areas of 18 individual 18-day-old Arabidopsis T_2 homozygous plants of Col-0 and ADAP-SRDX line were calculated using the region of interest (ROI) manager tool in ImageJ (Schneider et al. 2012). The leaf numbers were counted for leaves greater than 1 mm in length. All results were shown as the mean ± standard error (SE) of Col-0 and ADAP-SRDX plants from 18 individuals respectively. Statistical analysis between the two samples was performed using the two-tailed Student's t-test.

RNA extraction and cDNA synthesis

The RT-qPCR experiment was performed using ADAP-SRDX line with confirmed presence of SRDX transgene. RNA from 18-day-old Col-0 Arabidopsis and ADAP-SRDX plants were isolated with TRIzol reagent according to the manufacturer's instructions. The purity of RNA samples was quantified by A260:A280 ratios using the ND-1000 NanoDrop spectrophotometer (Thermo Scientific). First strand cDNA was synthesized using the Maxima First Strand cDNA Synthesis kit (Thermo Scientific).

RT-qPCR analysis

Reverse-transcription quantitative PCR (RT-qPCR) analysis was performed for *ADAP* and downstream GSL genes, *UDP-glycosyltransferase 74C1* (*UGT74C1*) and *isopropylmalate isomerase small subunit 1* (*IPMI1*) in three biological replicates. For qPCR analyses, the iTaq Universal SYBR® Green SuperMix kit (Bio-Rad) was used to prepare 10 µl qRT-PCR reactions containing 50 ng of first-strand cDNA and 300 nM of each primer pairs. Technical replicates and No Template Controls (NTCs) were run through the Eco48 Real-Time PCR system (PCRmax, UK). The protocol started with enzyme activation at 95 °C for 30 s, 40 PCR cycles (95 °C for 3 s and 60 °C for 30 s) followed by a melting curve analysis at 95 °C for 15 s, 55 °C for 15 s, and 95 °C for 15 s. The pre-installed Eco™ 48 Study Software was used to determine Ct values, and supports standard curve analysis and relative quantification experiments, which were used in this study. For standard curve analysis, five dilution points were used to ensure similar efficiency for each primer pair shown in Table S1. The relative gene expression levels were determined using *ACTIN2* as the housekeeping gene based on the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001). The significant difference of RT-qPCR data between two samples (Col-0 and ADAP-SRDX) was analyzed by t-test with significance at p-value < 0.05.

Results

Gene co-expression network analysis

In total, 102 known GSL genes from SuCComBase were used to build a GSL co-expression network of 752 genes and 9,121 edges. Fisher's exact test was performed to obtain GSL gene enrichment in each identified cluster. A total of 15 potential GSL genes were identified from eight significant clusters ($p < 0.05$).

Table 2 shows the list of potential GSL genes that were identified from highly significant clusters. Based on Table 2, ADAP is the only gene encoding TF and was selected for further studies and identified as a potential GSL gene. ADAP is found in the same cluster with four other known GSL genes, which are *MYC4, MYB28, MYB76*, and *AOP2,* as shown in Fig. 1.

Expression pattern analysis and identification of *cis*-regulatory elements of ADAP and known GSL genes

Plant hormones, such as jasmonic acid (JA), salicylic acid (SA), abscisic acid (ABA), and ethylene (ET) play roles in responses to abiotic and biotic stresses (Rehrig et al. 2014). Since GSLs are secondary metabolites that are potent in plant defense against insects and pests (Sønderby et al. 2010b), these conditions are selected to monitor the expression of *ADAP* and other known GSL genes. Figure 2 shows the *ADAP* expression, and several known Arabidopsis GSL genes across various conditions that include tissue development, tissue-specific (root, shoot, stem), abiotic stresses, and chemical stresses. *ADAP* and known GSL genes showed high expression under chemical stress conditions. Similar expression patterns of ADAP with 17 known GSL genes were identified from a total of 833 genes with an r-value cut-off range of 0.5–1 using Expression Angler (Fig. 2b). This information is useful to support the relationship between ADAP and the known GSL genes.

Specific *cis*-regulatory sequences in ADAP and UGT74C1 were identified using PlantPAN 3.0 as shown in Fig. 3. Several groups of transcription factor binding sites

Table 2 List of potential GSL genes from selected highly significant clusters

Cluster number	Cluster size	Number of potential genes	Potential genes	*p*-value
2	16	4	*JR1, TSA1, JAL23, BGLU18*	2.29E−08
4	13	3	*BGLU10, EMB1873, NLP1*	2.69E−07
6	12	4	*AT1G55450, ATNUDT7, RBOHD, IGMT4*	2.65E−05
51	6	1	*NIP6*	0.00022
52	6	1	*BGLU10*	0.00022
66	5	1	*CYP708A3*	0.001413
72	5	1	*SULTR2;1*	0.001413
77	5	1	*ADAP*	0.001413

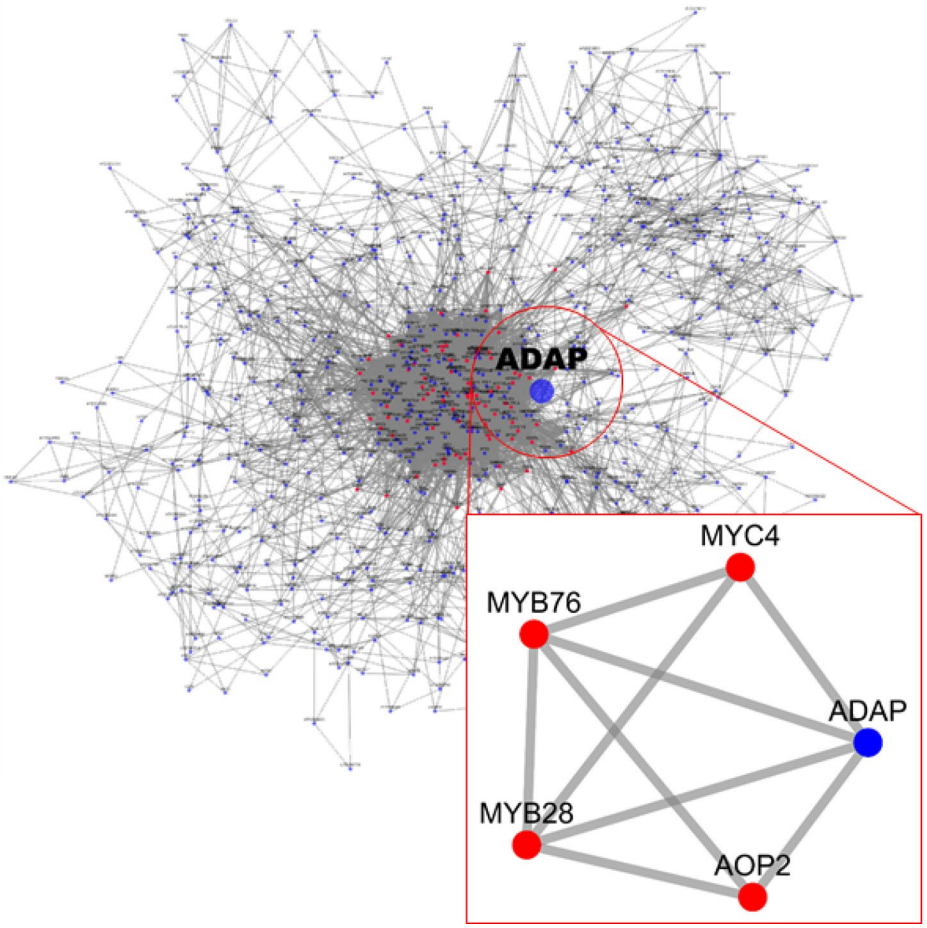

Fig. 1 A gene co-expression network that consists of 752 nodes and 9121 edges. Cluster 77 is incorporated in the gene network that consist of ADAP and known GSL genes (*MYC4*, *MYB28*, *MYB76*, and *AOP2*)

(TFbs) were identified from both ADAP and UGT74C1. Based on Fig. 3, *MYC2, MYC4,* and *OBP2* are the known GSL genes encoding TF that share similar TFbs with both target genes. More information on the identified TFbs is shown in Table 3.

Morphological comparisons of ADAP-SRDX and wild-type Col-0

Various bioinformatics analyses suggest the involvement of ADAP as a potential GSL gene in Arabidopsis. Thus,

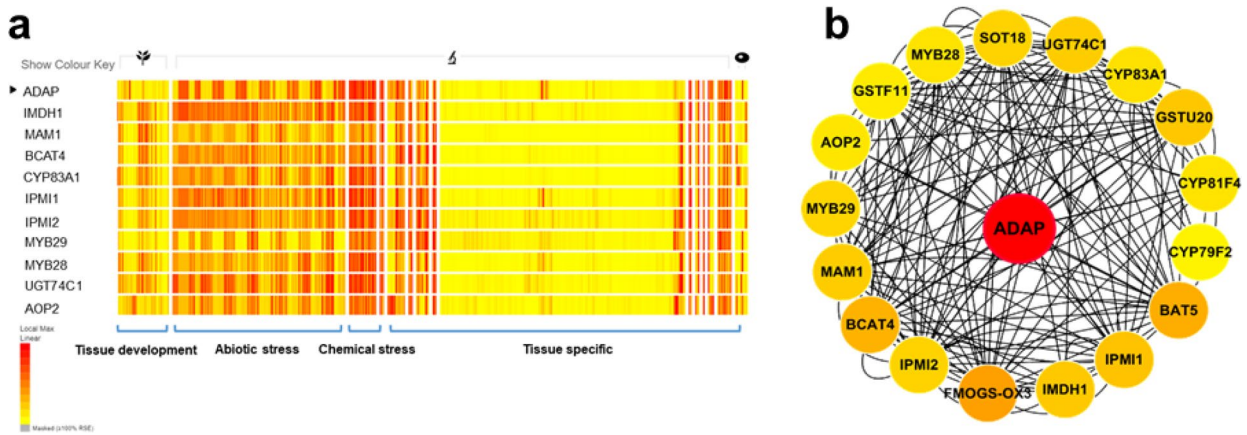

Fig. 2 Gene expression of AP2-like ethylene-responsive transcription factor (ADAP) has a similar expression pattern with several aliphatic GSL genes. **a** Expression pattern of *ADAP* across various conditions, including tissue development, abiotic and chemical stresses, and tissue-specific conditions. "Global" color gradient was selected to show similarities and differences in the expression levels of selected genes (red denotes high expression levels and yellow denotes low expression levels) **b** Expression pattern of *ADAP* with known GSL genes, such as *alkenyl hydroxalkyl producing 2* (*AOP2*), *hexahomomethionine N-hydroxylase* (*CYP79F2*), *UDP-glycosyltransferase 74C1* (*UGT74C1*) and *isopropylmalate isomerase small subunit 1* (*IPMI1*) in chemical stress

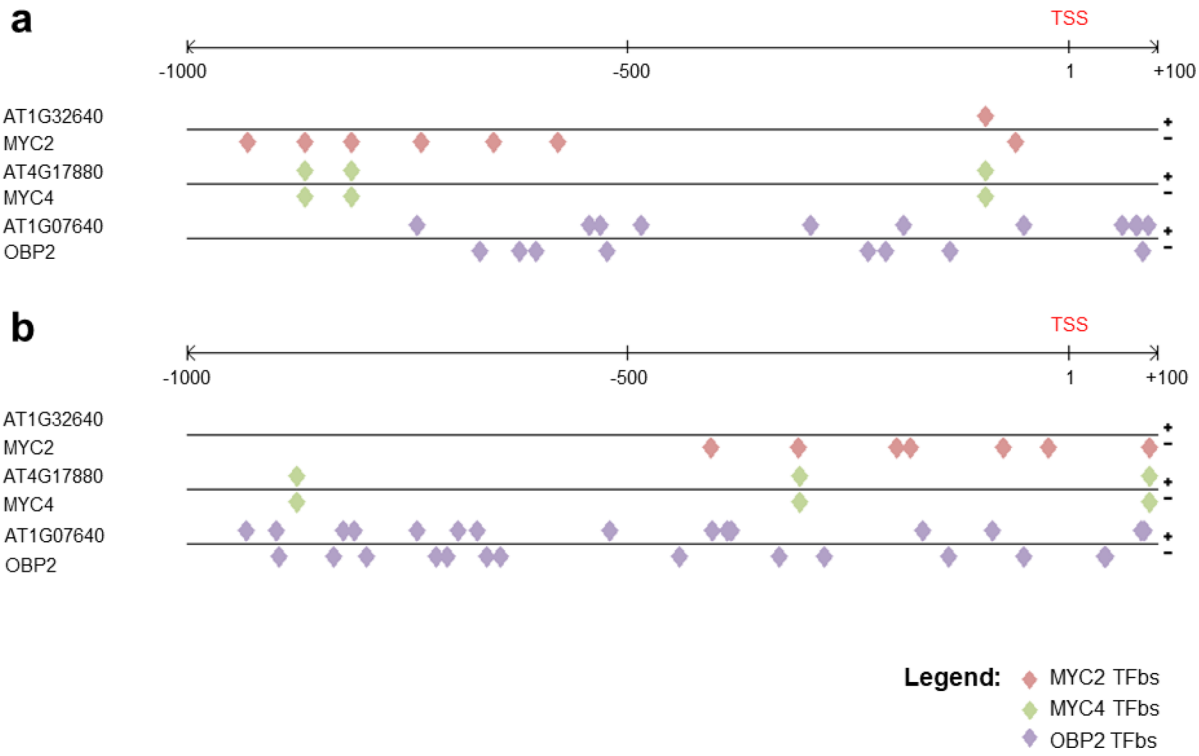

Fig. 3 Transcription factor binding sites (TFbs) of MYC2, MYC4, and OBP2 highlighted in **a** ADAP and **b** UGT74C1. *MYC2*, *MYC4,* and *OBP2* are the known GSL genes that share similar TFbs with both ADAP and UGT74C1. Each TFbs in ADAP and UGT74C1 sequences are shown with different colors

homozygous T_2 ADAP-SRDX plants were obtained and compared with the wild type Col-0. ADAP is a positive regulator of ABA response and seedling growth (Lee et al. 2009). The *ADAP-SRDX* plants showed significant ($p < 0.001$) differences in the leaf count (Fig. 4b) and size (Fig. 4c) compared to the Col-0. The *ADAP-SRDX* plants were approximately four times larger than the Col-0

Table 3 Transcription factor binding sites (TFbs) of OBP2, MYC2, and MYC4 in ADAP and UGT74C1

	OBP2			MYC2			MYC4		
Target gene	Position	Strand	Sequence	Position	Strand	Sequence	Position	Strand	Sequence
ADAP	− 739	+	AAGGG	− 813	−	cATGTG	− 866	+	CAAGTg
	− 668	−	ACCTT	− 95	+	CACATg	− 866	−	cAAGTG
	− 623	−	ACTTT	− 866	−	cAAGTG	− 813	+	CATGTg
	− 604	−	ACTTT	− 813	−	cATGTG	− 813	−	cATGTG
	− 543	+	AAAGT	− 95	+	CACATg	− 95	+	CACATg
	− 532	+	AAGGC	− 931	−	acaAGTAT	− 95	−	cACATG
	− 523	−	GCCTT	− 734	−	gcaGGTAT			
	− 485	+	AAAGA	− 652	−	actCGTAA			
	− 292	+	AAGGA	− 580	−	aaaCATAT			
	− 228	−	TCTTT	− 61	−	acaCGAAT			
	− 207	−	TCTTT						
	− 188	+	AAAGT						
	− 134	−	ACTTT						
	− 51	+	AAAGC						
	61	+	AAAGA						
	76	+	AAAGG						
	77	+	AAGGA						
	84	−	TCTTT						
	90	+	AAAGC						
UGT74C1	− 932	+	AAAGC	− 406	−	aaaAGTAT	− 875	+	GACGTg
	− 899	+	AAAGC	− 306	−	ccaCGTTT	− 875	−	gACGTG
	− 896	−	GCTTT	− 195	−	tcaCGCAT	− 305	+	CACGTt
	− 833	−	TCTTT	− 180	−	aaaCTTAT	-305	−	cACGTT
	− 823	+	AAAGG	− 75	−	acaAGTAT	92	+	CACGTa
	− 822	+	AAGGT	− 23	−	acaTGTAT	92	−	cACGTA
	− 810	+	AAGGA	91	−	tcaCGTAC			
	− 797	−	ACTTT						
	− 739	+	AAGGA						
	− 718	−	GCCTT						
	− 717	−	CCTTT						
	− 705	−	ACCTT						
	− 693	+	AAAGT						
	− 671	+	AAGGA						
	− 660	−	ACTTT						
	− 644	−	GCTTT						
	− 521	+	AAAGA						
	− 442	−	TCTTT						
	− 405	+	AAAGT						
	− 387	+	AAAGA						
	− 383	+	AAAGA						
	− 329	−	CCCTT						
	− 278	−	TCTTT						
	− 166	+	AAAGG						
	− 165	+	AAGGA						
	− 137	−	TCCTT						
	− 87	+	AAAGA						
	− 51	−	ACTTT						
	41	−	TCCTT						
	42	−	CCTTT						

Table 3 (continued)

Target gene	OBP2			MYC2			MYC4		
	Position	Strand	Sequence	Position	Strand	Sequence	Position	Strand	Sequence
	82	+	AAAGA						
	86	+	AAGGG						

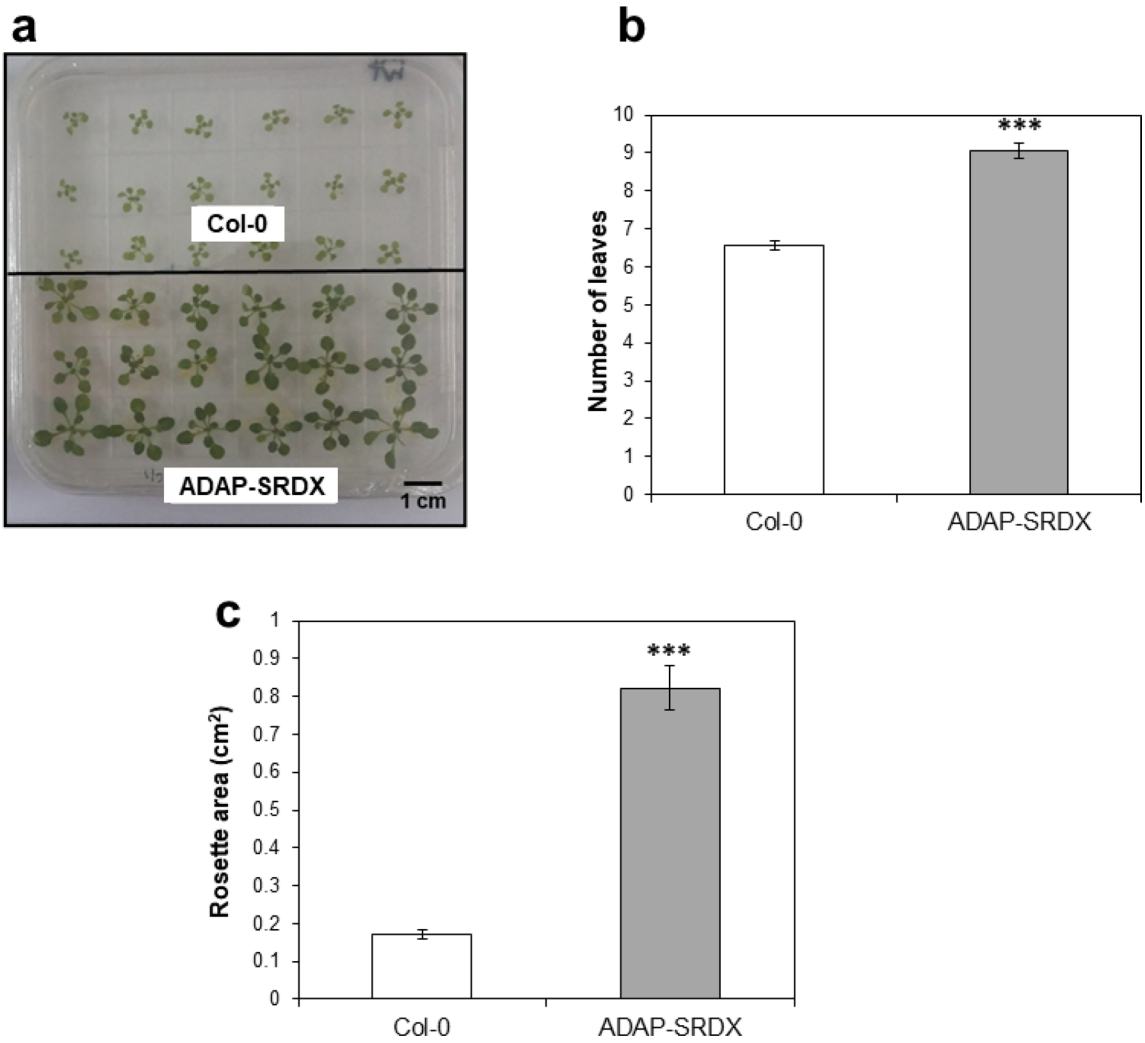

Fig. 4 Arabidopsis plants in Col-0 and ADAP-SRDX line. **a** Morphological comparisons between 18-days-old Arabidopsis Col-0 and ADAP-SRDX line. **b** The average leaf numbers of Col-0 and ADAP-SRDX plants. **c** The average rosette areas of Col-0 and ADAP-SRDX plants. Data are shown as mean ± standard error (SE) of the two samples from 18 individuals each. T-test ***$p < 0.001$

plants with an average rosette area of 0.82 ± 0.06 cm^2 and 0.17 ± 0.01 cm^2, respectively.

RT-qPCR analysis on the expression of ADAP and target genes

The expression of ADAP TF was monitored in two samples, namely Arabidopsis Col-0 and Arabidopsis ADAP-SRDX line (ADAP-SRDX) (Fig. 4a). Other downstream GSL genes sharing similar expression patterns with ADAP were selected (*UGT74C1* and *IPMI1*) to study their expression in these two samples. *ADAP* and *UGT74C1* were significantly overexpressed in the ADAP-SRDX line (Fig. 5). In the SRDX line, the expression of *ADAP, UGT74C1,* and *IPMI1* was induced up to 1900-, 35-, and 8-fold compared with the Col-0, respectively. The wide error bars reflect high biological variations between individual Arabidopsis plants (McIntyre et al. 2011).

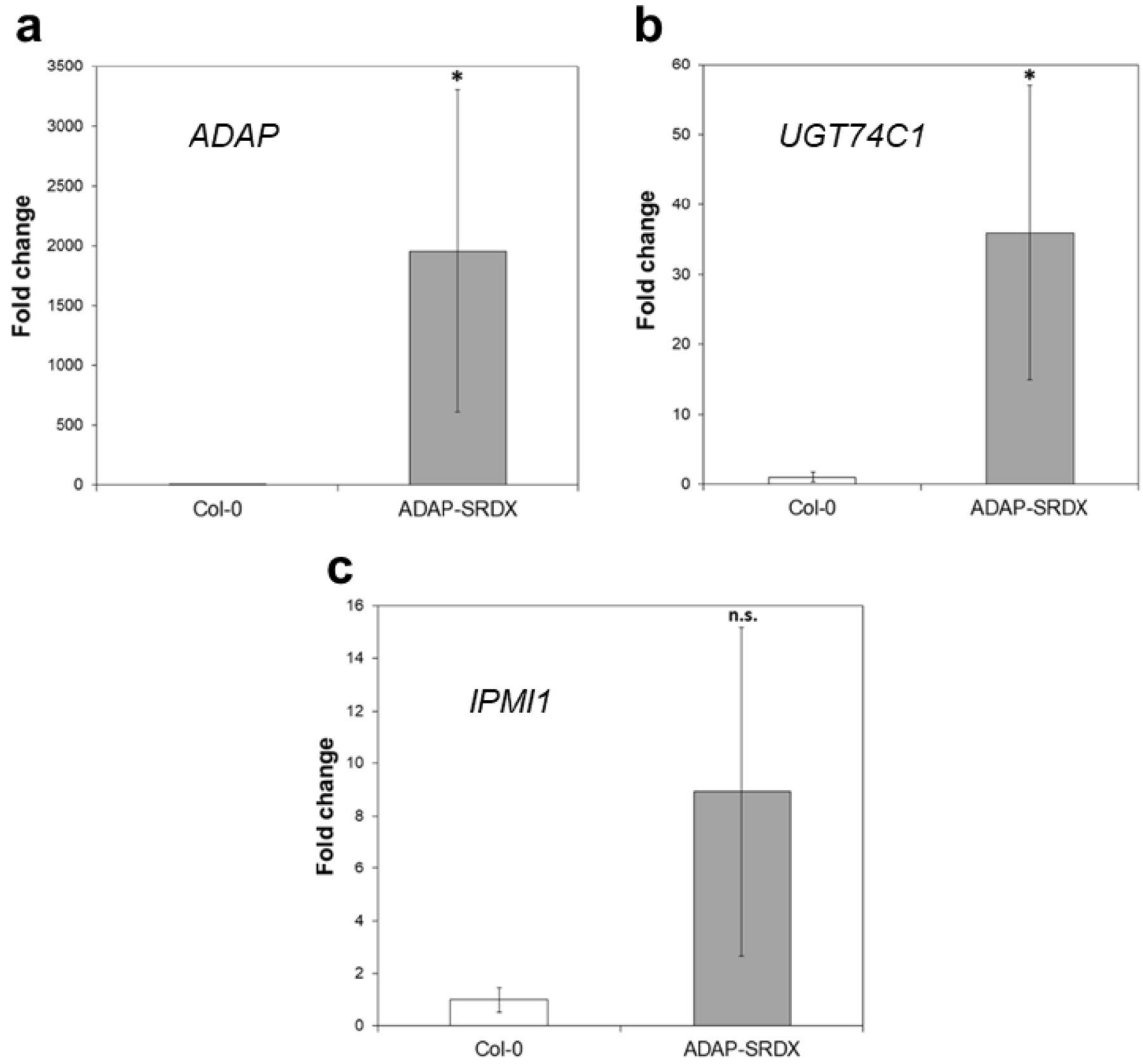

Fig. 5 ADAP gene expression in Col-0 and ADAP-SRDX line. **a** The expression of *ADAP* is significantly overexpressed in the ADAP-SRDX samples compared to the wild-type (Col-0). Expression pattern of downstream GSL genes, **b** *UDP-glycosyltransferase 74C1* (*UGT74C1*) and **c** *Isopropylmalate isomerase small subunit 1* (*IPMI1*) in Col-0 and ADAP-SRDX samples. The fold change represents relative gene expression levels between the two samples using the $2^{-\Delta\Delta Ct}$ method. Error bars represent SE of three biological replicates. T-test $*p < 0.05$; n.s.: not significant

Discussion

ADAP was identified as a potential GSL-related gene from comprehensive bioinformatics analyses, including gene co-expression network construction, graph clustering, and Fisher's exact test to identify the highly significant clusters. This assumption is made based on the 'guilt-by-association' principle, whereby *ADAP* was grouped in a highly significant cluster with known GSL genes such as *MYC4*, *MYB28*, *MYB76*, and *AOP2* (Fig. 1). *MYC4* (bHLH004) belongs to the basic helix-loop-helix (bHLH) family that is known to interact directly with the R2R3 domain MYB family. *MYB28* and *MYB76* TFs are members of the R2R3-MYB family that act as critical players in the modulation of aliphatic GSL genes (Gigolashvili et al. 2008; Hirai et al. 2007). The relationship between bHLH and R2R3-MYB families have been extensively described in which they interact and regulate the GSL biosynthesis that is important to overcome pest invasion. (Fernández-Calvo et al. 2011; Frerigmann 2016; Schweizer et al. 2013; Sønderby et al. 2010a). A recent study had analyzed the motif discovery that mediates the interaction of the MYB TFs (MYB28, MYB29, MYB76) and their bHLH partners (MYC2, MYC3, MYC4). Further analysis of the MYB-bHLH relationship uncovered a correlation in terms of affinity interaction and phenotype outcome that is regulated by the MYB-bHLH complex (Millard et al. 2019).

To further corroborate the involvement of ADAP TF in GSL biosynthesis, the pattern of gene expression between *ADAP* and known GSL genes was studied under chemical stress. Jasmonic acid (JA), salicylic acid (SA), abscisic acid (ABA), and ethylene (ET) effects on *Arabidopsis* Col-0 plants are among the chemical stress information that can be retrieved from Expression Angler. Figure 2b shows the co-expression module generated from the constructed GSL gene network between ADAP and known GSL genes which showed similar expression patterns under chemical stresses. The expression patterns in Expression Angler is calculated using Pearson correlation coefficient (PCC) shown by r-value. A common practice in gene co-expression network is identifying significant co-expressed genes by using the threshold value of r more than 0.7. However, an exception can be applied when some variances would cause noise in the dataset (Usadel et al. 2009). Furthermore, in a recent study, any edge with PCC < 0.3 used to construct several co-expression modules in Arabidopsis was discarded (Wisecaver et al. 2017). Thus, 0.5 was selected as a threshold value in this study to generate as many interacting partners between *ADAP* and known GSL genes.

The *ADAP* gene co-expression network in Fig. 2b comprises a majority of known aliphatic GSL genes that have specific roles, which include TFs (*MYB28, MYB29*), protein transporter (BAT5) as well as enzymes that are involved in the side-chain elongation (*IMDH1, IPMI2 MAM1, BCAT4*), core structure synthesis (*CYP79F2, CYP83A1, GSTU20, GSTF11, UGT74C1, FMOGS-OX3*) and the side-chain modification (*AOP2*). In order to verify the function of *ADAP* in GSL biosynthesis, two known downstream aliphatic GSL genes (*UGT74C1*, *IPMI1*) were selected and the expression of these genes was studied in Arabidopsis ADAP-SRDX and Col-0 samples. By using the CRES-T approach, the coding region of ADAP fused at the C-terminal domain with a repressor domain called the plant-specific EAR motif repressor domain (SRDX). The fusion of the SRDX domain with a TF would create a dominant ADAP suppressor that suppresses the expression of downstream target genes (Mitsuda and Ohme-Takagi 2009; Mitsuda et al. 2011). Therefore, it is suggested that the fusion of ADAP TF with the SRDX domain would suppress the expression of target GSL genes, which are *UGT74C1* and *IPMI1*. Furthermore, the phenotypes discovered from the transgenic SRDX plants would show similar observation seen in the loss-of-function mutants of the gene that encodes for the respective TF (Mitsuda et al. 2011).

ADAP is a member of the AP2/ERF TF family that regulates plant defense against necrotrophic pathogens involving jasmonate (JA) and ethylene (ET) signaling pathways (Broekgaarden et al. 2015; Huang et al. 2015). Furthermore, ADAP is known as a positive regulator of abscisic acid (ABA) response that plays essential roles in plant responses to stress (Chen et al. 2020; Lee et al. 2009). The presence of ABA, which can influence the biosynthesis of GSL in cabbage sprouts, was confirmed in a previous study. In the analysis, treatment with ABA significantly increased the content of GSL, the formation of isothiocyanate and the myrosinase activity of the 5-day-old cabbage sprouts. These findings indicated that ABA played a key role in GSL biosynthesis and metabolism in the edible parts of cabbage sprouts (Wang et al. 2015).

ADAP is subdivided into AP2 subgroup that contains two AP2 domains. The subgroup are known to control plant growth during several developmental stages (Heyman et al. 2018). Thus, the observation of Arabidopsis ADAP-SRDX plants (Fig. 3) supported the capability of the CRES-T approach to validate the involvement of ADAP in plant development. Hypothetically, the transgenic ADAP line should create a dominant ADAP suppressor that suppresses the expression of target GSL genes. However, RT-qPCR analysis showed that the ADAP-SRDX line showed overexpression of *ADAP* and its predicted target GSL genes (*UGT74C1* and *IPMI1*) approximately 1900-, 35-, and 8-fold higher than the Col-0, respectively. Thus, ADAP may act as a negative regulator in the aliphatic GSL biosynthesis where it was significantly over-expressed via a feedback mechanism that also increase the expression of downstream aliphatic GSL genes (*UGT74C1* and *IPMI1*). UGT74C1 was previously proposed in the aliphatic GSL biosynthesis from the Arabidopsis genome and transcriptome data supports its co-expression with other known Met-derived GSLs (Gachon et al. 2005). Further analysis corroborates those findings through phylogenetic, gene expression, and knockout study on *ugt74c1-2* (Grubb et al. 2014).

Transcription factors (TFs) are known as sequence-specific DNA binding proteins that act as regulators of gene expression through the recognition of specific *cis*-regulatory sequences in promoter regions. Based on the *cis*-regulatory sequence analysis between *ADAP* and *UGT74C1*, the bHLH TFs (*MYC2* and *MYC4*) and *OBP2* are the known GSL regulatory genes that share similar TF binding sites (TFbs) with both genes (Fig. 3). The analysis identified potential TF recognition sites within the promoter and potential candidate genes involved in the GSL regulatory mechanism. MYC2 and MYC4 were previously described as positive regulators in aliphatic and indolic GSL biosynthesis (Fernández-Calvo et al. 2011; Li et al. 2018; Schweizer et al. 2013). Specifically, MYC2 reacts to jasmonate secretion and regulates MYB TFs in GSL biosynthesis (Bell 2019; Chini et al. 2016).

Previously, Skirycz et al. (2006) confirmed the involvement of OBP2 with CYP83B1 which later suggested its involvement in the regulatory network that regulates indolic GSL biosynthesis in Arabidopsis. OBP2 is a DOF (DNA binding with one finger) transcription factor. Its expression

was induced in response to herbivores, and by methyl jasmonate treatment. Both treatments would trigger indolic GSL accumulation. Specifically, the overexpression of OBP2 activates *CYP83B1* expression, and RNA interference-mediated OBP2 blockade leads to reduced expression of *CYP83B1*. Collectively, these data provide evidence that OBP2 is involved in the regulatory network of GSL biosynthesis in Arabidopsis (Skirycz et al. 2006). Based on this information, we propose the regulatory mechanism of ADAP as shown in Fig. 6, which shares the same TFbs with regulatory genes, MYC2, MYC3, and OBP2, as well as the GSL biosynthetic gene, UGT74C1.

Conclusion

ADAP is known as a positive regulator of ABA that plays a vital role in stress response in plants. ABA has shown to affect the GSL biosynthesis in *Brassicaceae* plants by increasing the GSL content and its derivative (isothiocyanate). Based on several bioinformatics analyses, we propose ADAP as a possible negative regulator in aliphatic GSL biosynthesis that was validated by the RT-qPCR analysis of *ADAP-SRDX* line. Significant overexpression of downstream GSL genes such as *UGT74C1* and *IPMI1* were found in the ADAP-SRDX line suggesting their role as essential key players in glucosinolate biosynthesis. This study shows the preliminary findings on the capability of ADAP to regulate aliphatic GSL biosynthesis. Furthermore, the substantial significant over-expression of *ADAP* gene in the ADAP-SRDX line also suggests the TF activity that has a negative effect on *UGT74C1* and *IPMI1* expression via a feedback mechanism in Arabidopsis. However, further experimental validation, such as a targeted metabolomics approach and functional analysis of downstream genes is required to confirm the regulatory mechanism of ADAP in GSL biosynthesis as proposed in this work.

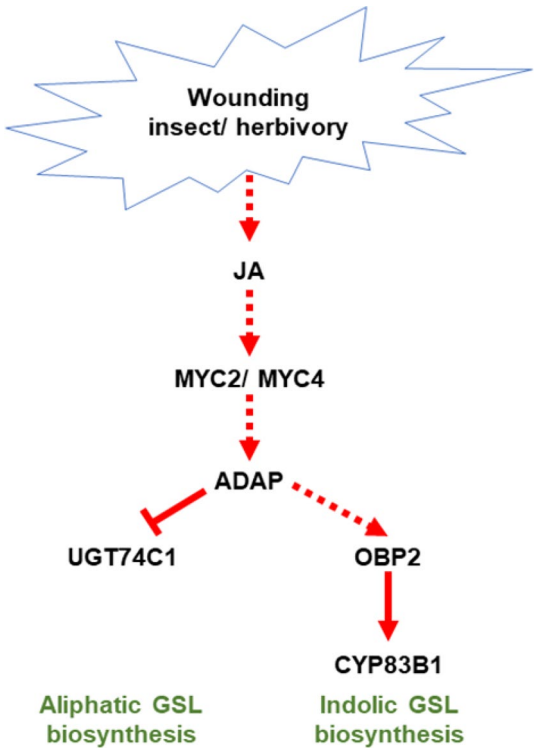

Fig. 6 Proposed regulatory mechanism of ADAP and UGT74C1 that have similar TFbs with known GSL genes TF, MYC2, MYC4, and OBP2. OBP2 acts as a positive regulator in the indolic GSL biosynthesis. Based on the qPCR analysis, ADAP may also involve in the aliphatic GSL biosynthesis based on the significant upregulation of UGT74C1 in the ADAP-SRDX line

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10265-021-01257-9.

Acknowledgements This work was supported by ERGS/1/2013/STG07/UKM/02/3 and UKM-MGI-NBD0005-2007 awarded to ZAMH by the Malaysian Ministry of Higher Education and Ministry of Science, Technology and Innovation.

Compliance with ethical standards

Conflict of interest None declared.

References

Altaf-Ul-Amin M, Shinbo Y, Mihara K et al (2006) Development and implementation of an algorithm for detection of protein complexes in large interaction networks. BMC Bioinform 7:207. https://doi.org/10.1186/1471-2105-7-207

Altaf-Ul-Amin M, Wada M, Kanaya S (2012) Partitioning a PPI network into overlapping modules constrained by high-density and periphery tracking. ISRN Biomath 2012:1–11. https://doi.org/10.5402/2012/726429

Aoki Y, Okamura Y, Tadaka S et al (2016) ATTED-II in 2016: a plant coexpression database towards special online collection. Plant Cell Physiol 57:1–9. https://doi.org/10.1093/pcp/pcv165

Bell L (2019) The biosynthesis of glucosinolates: insights, inconsistencies, and unknowns. Annu Plant Rev 2:1–31. https://doi.org/10.1002/9781119312994.apr0708

Broekgaarden C, Caarls L, Vos IA et al (2015) Ethylene: traffic controller on hormonal crossroads to defense. Plant Physiol 169:2371–2379. https://doi.org/10.1104/pp.15.01020

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. Phytochemistry 62:471–481

Cen H, Ye W, Liu Y et al (2016) Overexpression of a chimeric gene, OsDST-SRDX, improved salt tolerance of perennial ryegrass. Sci Rep 6:1–13. https://doi.org/10.1038/srep27320

Chen K, Li GJ, Bressan RA et al (2020) Abscisic acid dynamics, signaling, and functions in plants. J Integr Plant Biol 62:25–54. https://doi.org/10.1111/jipb.12899

Chini A, Gimenez-Ibanez S, Goossens A, Solano R (2016) Redundancy and specificity in jasmonate signalling. Curr Opin Plant Biol 33:147–156

Choi H, Cho M, Gu H, Yoon D (2010) Indole-3-carbinol induces apoptosis through p53 and activation of caspase-8 pathway in lung cancer A549 cells. Food Chem Toxicol 48:883–890. https://doi.org/10.1016/j.fct.2009.12.028

Choi SH, Park S, Lim YP et al (2014) Metabolite profiles of glucosinolates in cabbage varieties (*Brassica oleracea* var. *capitata*) by season, color, and tissue position. Hortic Environ Biotechnol 55:237–247. https://doi.org/10.1007/s13580-014-0009-6

Chung IM, Rekha K, Rajakumar G, Thiruvengadam M (2016) Production of glucosinolates, phenolic compounds and associated gene expression profiles of hairy root cultures in turnip (*Brassica rapa* ssp *rapa*). 3 Biotech 6:1–16. https://doi.org/10.1007/s13205-016-0492-9

Clay NK, Adio AM, Denoux C et al (2009) Glucosinolate metabolites required for an Arabidopsis innate immune response. Science (80-) 323:95–101

Fernández-Calvo P, Chini A, Fernández-Barbero G et al (2011) The Arabidopsis bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. Plant Cell 23:701–715. https://doi.org/10.1105/tpc.110.080788

Fisher RA (1992) Statistical methods for research workers. In: Samuel K, Norman LJ (eds) Breakthroughs in statistics, vol 2. Springer, Berlin, pp 66–70

Frerigmann H (2016) Glucosinolate regulation in a complex relationship—MYC and MYB—no one can act without each other. Adv Bot Res 80:57–97. https://doi.org/10.1016/bs.abr.2016.06.005

Frerigmann H, Gigolashvili T (2014) Indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. Mol Plant 7:814–828. https://doi.org/10.1093/mp/ssu004

Gachon CMM, Langlois-Meurinne M, Henry Y, Saindrenan P (2005) Transcriptional co-regulation of secondary metabolism enzymes in Arabidopsis: functional and evolutionary implications. Plant Mol Biol 58:229–245

Geu-Flores F, Nielsen MT, Nafisi M et al (2009) Glucosinolate engineering identifies a γ-glutamyl peptidase. Nat Chem Biol 5:575–577. https://doi.org/10.1038/nchembio.185

Gigolashvili T, Berger B, Mock H, Mu C (2007) The transcription factor HIG1 / MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. Plant J 50:886–901. https://doi.org/10.1111/j.1365-313X.2007.03099.x

Gigolashvili T, Engqvist M, Yatusevich R et al (2008) HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. New Phytol 177:627–642. https://doi.org/10.1111/j.1469-8137.2007.02295.x

Gillis J, Pavlidis P (2012) "Guilt by association" is the exception rather than the rule in gene networks. PLoS Comput Biol. https://doi.org/10.1371/journal.pcbi.1002444

Grubb CD, Zipp BJ, Ludwig-Muller J et al (2004) Arabidopsis glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. Plant J 40:893–908. https://doi.org/10.1111/j.1365-313X.2004.02261.x

Grubb CD, Zipp BJ, Kopycki J et al (2014) Comparative analysis of Arabidopsis UGT74 glucosyltransferases reveals a special role of UGT74C1 in glucosinolate biosynthesis. Plant J. https://doi.org/10.1111/tpj.12541

Harun S, Abdullah-Zawawi M-R, A-Rahman MRA et al (2019) SuCComBase: a manually curated repository of plant sulfur-containing compounds. Database (Oxford). https://doi.org/10.1093/database/baz021

Harun S, Abdullah-Zawawi MR, Goh HH, Mohamed-Hussein ZA (2020) A comprehensive gene inventory for glucosinolate biosynthetic pathway in *Arabidopsis thaliana*. J Agric Food Chem 68:7281–7297. https://doi.org/10.1021/acs.jafc.0c01916

Heyman J, Canher B, Bisht A et al (2018) Emerging role of the plant ERF transcription factors in coordinating wound defense responses and repair. J Cell Sci. https://doi.org/10.1242/jcs.208215

Hirai MY, Sugiyama K, Sawada Y et al (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proc Natl Acad Sci 104:6478–6483

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

Huang P, Catinot J, Zimmerli L (2015) Ethylene response factors in Arabidopsis immunity. J Exp Bot 13040:2–11. https://doi.org/10.1093/jxb/erv518

Kanehisa M, Sato Y, Kawashima M et al (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D462. https://doi.org/10.1093/nar/gkv1070

Kazama D, Itakura M, Kurusu T et al (2013) Identification of chimeric repressors that confer salt and osmotic stress tolerance in Arabidopsis. Plants 2:769–785. https://doi.org/10.3390/plants2040769

Klein M, Papenbrock J (2009) Kinetics and substrate specificities of desulfo-glucosinolate sulfotransferases in *Arabidopsis thaliana*. Physiol Plant 135:140–149. https://doi.org/10.1111/j.1399-3054.2008.01182.x

Knill T, Schuster J, Reichelt M et al (2008) Arabidopsis branched-chain aminotransferase 3 functions in both amino acid and glucosinolate biosynthesis. Plant Physiol 146:1028–1039. https://doi.org/10.1104/pp.107.111609

Lee SJ, Cho DI, Kang JY, Kim SY (2009) An ARIA-interacting AP2 domain protein is a novel component of ABA signaling. Mol Cells 27:409–416. https://doi.org/10.1007/s10059-009-0058-3

Lee T, Yang S, Kim E et al (2014) AraNet v2: an improved database of co-functional gene networks for the study of Arabidopsis thaliana and 27 other nonmodel plant species. Nucleic Acids Res. https://doi.org/10.1093/nar/gku1053

Li B, Tang M, Nelson A et al (2018) Network-guided discovery of extensive epistasis between transcription factors involved in aliphatic glucosinolate biosynthesis. Plant Cell 30:178–195. https://doi.org/10.1105/tpc.17.00805

Liesecke F, Daudu D, De Bernonville RD et al (2018) Ranking genome-wide correlation measurements improves microarray and RNA-seq based global and targeted co-expression networks. Sci Rep 8:1–16. https://doi.org/10.1038/s41598-018-29077-3

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25:402–408. https://doi.org/10.1006/meth.2001.1262

McIntyre LM, Lopiano KK, Morse AM et al (2011) RNA-seq: Technical variability and sampling. BMC Genom 12:293. https://doi.org/10.1186/1471-2164-12-293

Mikkelsen MD, Naur P, Halkier BA (2004) Arabidopsis mutants in the C–S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. Plant J 37:770–777. https://doi.org/10.1111/j.1365-313X.2004.02002.x

Millard PS, Weber K, Kragelund BB, Burow M (2019) Specificity of MYB interactions relies on motifs in ordered and disordered contexts. Nucleic Acids Res 47:9592–9608. https://doi.org/10.1093/nar/gkz691

Miranda Rossetto MR, Shiga TM, Vianello F, Pereira Lima GP (2013) Analysis of total glucosinolates and chromatographically purified benzylglucosinolate in organic and conventional vegetables. LWT Food Sci Technol 50:247–252. https://doi.org/10.1016/j.lwt.2012.05.022

Mitsuda N, Ohme-Takagi M (2009) Functional analysis of transcription factors in Arabidopsis. Plant Cell Physiol 50:1232–1248. https://doi.org/10.1093/pcp/pcp075

Mitsuda N, Matsui K, Ikeda M, et al (2011) CRES-T, an effective gene silencing system utilizing chimeric repressors. In: Clifton NJ (ed) Methods in molecular biology. Springer International Publishing, Berlin, pp 87–105

Montojo J, Zuberi K, Rodriguez H et al (2014) GeneMANIA: fast gene network construction and function prediction for cytoscape. F1000Research 153:1–7. https://doi.org/10.12688/f1000research.4572.1

Mueller LA, Zhang P, Rhee SY (2003) AraCyc: a biochemical pathway database for Arabidopsis. Plant Physiol 132:453–460. https://doi.org/10.1104/pp.102.017236.ual

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant 15:473–497. https://doi.org/10.1111/j.1399-3054.1962.tb08052.x

Naur P, Petersen BL, Mikkelsen MD et al (2003) CYP83A1 and CYP83B1, two nonredundant cytochrome P450 enzymes metabolizing oximes in the biosynthesis of glucosinolates in Arabidopsis. Plant Physiol 133:63–72. https://doi.org/10.1104/pp.102.019240.1

Padilla G, Cartea ME, Velasco P et al (2007) Variation of glucosinolates in vegetable crops of *Brassica rapa*. Phytochemistry 68:536–545

Rao CV (2013) Benzyl isothiocyanate: double trouble for breast cancer cells. Cancer Prev Res 6:760–763. https://doi.org/10.1158/1940-6207.CAPR-13-0242

Rehrig EM, Appel HM, Jones AD et al (2014) Roles for jasmonate- and ethylene-induced transcription factors in the ability of Arabidopsis to respond differentially to damage caused by two insect herbivores. Front Plant Sci 5:1–14. https://doi.org/10.3389/fpls.2014.00407

Sawada Y, Kuwahara A, Nagano M et al (2009) Omics-based approaches to methionine side chain elongation in Arabidopsis: characterization of the genes encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase. Plant Cell Physiol 50:1181–1190. https://doi.org/10.1093/pcp/pcp079

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675. https://doi.org/10.1038/nmeth.2089

Schweizer F, Fernández-calvo P, Zander M et al (2013) Arabidopsis basic helix-loop-helix transcription factors insect performance, and feeding behavior. Plant Cell 25:3117–3132. https://doi.org/10.1105/tpc.113.115139

Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. https://doi.org/10.1101/gr.1239303.metabolite

Skirycz A, Reichelt M, Burow M et al (2006) DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in Arabidopsis. Plant J 47:10–24. https://doi.org/10.1111/j.1365-313X.2006.02767.x

Sønderby IE, Hansen BG, Bjarnholt N et al (2007) A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. PLoS ONE e1322:1–16. https://doi.org/10.1371/journal.pone.0001322

Sønderby IE, Burow M, Rowe HC et al (2010a) A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in Arabidopsis. Plant Physiol 153:348–363. https://doi.org/10.1104/pp.109.149286

Sønderby IE, Geu-flores F, Halkier BA (2010b) Biosynthesis of glucosinolates—gene discovery and beyond. Trends Plant Sci 15:283–290. https://doi.org/10.1016/j.tplants.2010.02.005

Szklarczyk D, Gable AL, Lyon D et al (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47:D607–D613. https://doi.org/10.1093/nar/gky1131

Usadel B, Obayashi T, Mutwil M et al (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell Environ 32:1633–1651. https://doi.org/10.1111/j.1365-3040.2009.02040.x

Waese J, Fan J, Pasha A et al (2017) ePlant: Visualizing and exploring multiple levels of data for hypothesis generation in plant biology. Plant Cell 29:1806–1821. https://doi.org/10.1105/tpc.17.00073

Wang Z, Yang R, Guo L et al (2015) Effects of abscisic acid on glucosinolate content, isothiocyanate formation and myrosinase activity in cabbage sprouts. Int J Food Sci Technol 50:1839–1846. https://doi.org/10.1111/ijfs.12848

Warde-Farley D, Donaldson SL, Comes O et al (2010) The GeneMANIA prediction server : biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 38:W214–W220. https://doi.org/10.1093/nar/gkq537

Wisecaver JH, Borowsky AT, Tzin V et al (2017) A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. Plant Cell 29:944–959. https://doi.org/10.1105/tpc.17.00009

Wittstock U, Halkier BA (2000) Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* L. catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. J Biol Chem 275:14659–14666

Wong DCJ (2020) Network aggregation improves gene function prediction of grapevine gene co-expression networks. Plant Mol Biol 103:425–441. https://doi.org/10.1007/s11103-020-01001-2