A novel aldo-keto reductase gene is involved in 6′-deoxychalcone biosynthesis in dahlia (*Dahlia variabilis*)

Authors:

Sho Ohno[1]*, Haruka Yamada[1], Kei Maruyama[1], Ayumi Deguchi[1]**, Yasunari Kato[1], Mizuki Yokota[1], Fumi Tatsuzawa[2], Munetaka Hosokawa[1]***, Motoaki Doi[1]

Affiliations:

1. Graduate School of Agriculture, Kyoto University, Kyoto, Kyoto, 606-8502, Japan

2. Faculty of Agriculture, Iwate University, Morioka, Iwate, 020-8550, Japan

**Current address: Chiba University, Chiba, 271-8510, Japan

***Current address: Kindai University, Nara, 631-0052, Japan

*Corresponding author: Sho Ohno

Email: ohno.sho.3c@kyoto-u.ac.jp

ORCID: orcid.org/0000-0002-0810-0327

Main conclusion: A novel gene belonging to the aldo-keto reductase 13 family is involved in isoliquiritigenin biosynthesis in dahlia.

Funding

This study was supported by the Grant-in-Aid for Young Scientists (No. 18K14455) from the Japan Society for the Promotion of Science to SO.

Author Contribution statement

SO conceived the study. SO, MH and MD designed the experiments. SO, HY, KM, AD, YK, MY, and FT conducted the experiments; SO wrote the manuscript. All the authors read and approved the manuscript.

Abstract

The yellow pigments of dahlia flowers are derived from 6'-deoxychalcones, which are synthesized via a two-step process, involving the conversion of 3-malonyl-CoA and 4-coumaloyl-CoA into isoliquiritigenin in the first step, and the subsequent generation of butein from isoliquiritigenin. The first step reaction is catalyzed by chalcone synthase (CHS) and aldo-keto reductase (AKR). AKR has been implicated in the isoflavone biosynthesis in legumes, however, isolation of butein biosynthesis related AKR members are yet to be reported. A comparative RNA-seq analysis between two dahlia cultivars, "Shukuhai" and its butein-deficient lateral mutant "Rinka", was used in this study to identify a novel *AKR* gene involved in 6'-deoxychalcone biosynthesis. *DvAKR1* encoded a AKR 13 sub-family protein with significant differential expression levels, and was phylogenetically distinct from the chalcone reductases, which belongs to the AKR 4A sub-family in legumes. DNA sequence variation and expression profiles of *DvAKR1* gene were correlated with 6'-deoxychalcone accumulation in the tested dahlia cultivars. A single over-expression analysis of *DvAKR1* was not sufficient to initiate the accumulation of isoliquiritigenin in tobacco, in contrast, its co-overexpression with a *chalcone 4'-O-glucosyltransferase* (*Am4'CGT*) from *Antirrhinum majus* and a MYB transcription factor, *CaMYBA* from *Capsicum annuum* successfully induced isoliquiritigenin accumulation. In addition, *DvAKR1* homologous gene expression was detected in Coreopsideae species accumulating 6'-deoxychalcone, but not in Asteraceae species lacking 6'-deoxychalcone production. These results not only demonstrate the involvement of *DvAKR1* in the

biosynthesis of 6'-deoxychalcone in dahlia, but also show that 6'-deoxychalcone occurrence in Coreopsideae species developed evolutionarily independent from legume species.

Key words: AKR, butein, chalcone reductase, flavonoids, isoliquiritigenin, yellow pigment

Abbreviations

4′CGT	chalcone 4′-*O*-glucosyltransferase
AKR	aldo-keto reductase
AS	aureusidin synthase
AUS	aurone synthase
CH3H	chalcone 3-hydroxylase
CHR	chalcone reductase
CHS	chalcone synthase
HPLC	high-performance liquid chromatography

Introduction

Flower color is one of the most important traits in ornamental plants, and a key driver of floriculture industry. Yellow coloration is mainly derived from the accumulation of carotenoids or betaxanthins, and their absence in other plant species cause lack of bright yellow flowers. Some flavonoid compounds are known to regulate the yellow flower

coloration, including butein (2′,3,4,4′-tetrahydroxychalcone), which confers bright yellow color in flowers, and has only been detected in limited plant species, such as *Dahlia variabilis* (Price 1939), *Cosmos sulphureus* (Geissman 1942), and *Coreopsis grandiflora* (Geissman and Heaton 1943). Butein is presumably synthesized from common flavonoid substrates, including 3-malonyl-CoA and 4-coumaroyl-CoA, which are converted to 2′,4,4′-trihydroxychalcone (isoliquiritigenin) by chalcone synthase (CHS), and chalcone reductase (CHR), an aldo-keto reductase (AKR) superfamily gene, then hydroxylated by chalcone 3-hydroxylase (CH3H) (Fig. 1). A gene encoding CH3H has previously been characterized in *C. sulphureus* (Schlangen et al. 2010b), while an allelic variant of flavonoid 3′-hydroxylase (F3′H) in dahlia was shown to exhibit a CH3H activity (Schlangen et al. 2010a). Aurone is another yellow flavonoid compound, which is found in snapdragon (*A. majus*) and Asteraceae species (Miosic et al. 2013). In snapdragon, 4-hydroxyaurones are synthesized by two key enzymes, including aureusidin synthase (AmAS1) and Am4′CGT, whereas their synthesis from 6'-deoxychalcone in the genus *Coreopsis* is catalyzed by aurone synthase (AUS) (Fig. 1) (Kaintz et al. 2014; Molitor et al. 2015). Interestingly, the gene catalyzing isoliquiritigenin biosynthesis in the 6'-deoxychalcone and 4-deoxyaurone biosynthetic pathway is yet to be identified.

Dahlias are autoallooctoploid (2n = 8x = 64), and popular ornamental plants in the Asteraceae family (Gatt et al. 1998). The plants exhibit wide variations in flower color due to flavonoid and related pigments. For example, numerous genes associated with flavonoid biosynthesis have been characterized in their ray florets (Suzuki et al. 2002; Ohno et al. 2011a, 2011b, 2013b, 2018a; Deguchi et al. 2013). Despite producing flavonoids with no yellow carotenoids or betaxanthins in its ray florets, dahlias cultivars develop numerous bright yellow flowers. Butein was first identified as the yellow flower pigment in dahlia by Price (1939). Later, characterization of butein structures, including

4′-malonylsopholoside and 4′-malonylglucoside was reported by Harborne et al. (1990). Subsequently, our previous study identified five isoliquiritigenin derivatives and five butein derivatives from ‘Shukuhai’ (Fig. 2a) ray florets using fast-atom-bombardment mass spectrometry and nuclear magnetic resonance techniques. Of the derivatives, butein 4′-malonylglucoside showed the highest correlation coefficient with the international commission on illumination (CIE) color space value of *b** in the ray florets of nine yellow flower cultivars (Ohno et al. 2021). Butein interacts with anthocyanin to generate orange to red floral coloration (Ohno et al. 2013a), and its derivatives have also been detected in the leaves (Ohno et al. 2018b). In dahlia, butein always co-accumulate with isoliquiritigenin in ray florets, suggesting that isoliquiritigenin biosynthesis is crucial for 6'-deoxychalcone biosynthesis (Ohno et al. 2013a).

CHR (although named chalcone reductase, it is most likely that its substantial substrate is not chalcone, but a polyketide intermediate produced during chalcone synthase catalysis: Bomati et al. 2005) has only been characterized in the isoflavone synthetic pathway in legumes. NAD(P)H-dependent 6'-deoxychalcone synthase activity was initially reported in licorice (*Glycyrrhiza echinate*) and soybean (*Glycine max*) (Ayabe et al. 1988; Welle and Grisebach 1988), and the cDNA sequence of *GmCHR1* was first characterized in soybean (Welle et al. 1991). All *GmCHR* orthologs in leguminous plants associated with isoliquiritigenin biosynthesis in the isoflavone biosynthetic pathway belongs to the AKR 4A sub-family (Jez et al. 1997; Bomati et al. 2005). However, no orthologous AKR 4A sub-family gene was found by preliminary RNA-seq analysis in dahlia, suggesting that its 6'-deoxychalcone biosynthesis is regulated by different gene(s) from those of Leguminosae. Thus, to explore the candidate genes involved in 6'-deoxychalcone biosynthesis in dahlia, a red-white cultivar ‘Shukuhai’ (SH) and its two-lateral mutant series, ‘Iwaibune’ (IB) (dark red-white) and ‘Rinka’ (RK) (purple-white),

were used in this study (Fig. 2a). ‘SH’ and ‘IB’ accumulates anthocyanins, flavones, and 6'-deoxychalcones in its red tissues, while ‘RK’ accumulates anthocyanins and flavones, but not 6'-deoxychalcones in its purple tissues. Using comparative transcriptome analysis, we identified a candidate contig designated, *DvAKR1*, belonging to the AKR 13 sub-family, and subsequently performed its functional verification using transgenic approaches.

Materials and methods

Plant materials

The red-white bicolor dahlia cultivar ‘SH’ and its two-lateral mutant lines ‘IB’ and ‘RK’ were used. ‘IB’ is a lateral mutant of ‘SH’ and it has dark red-white ray florets, and ‘RK’ is a lateral mutant of ‘IB’ and it has purple-white ray florets (Fig. 2a). Ray florets for subsequent analyses were collected at different developmental stages (Fig. 2b) from field- or greenhouse-grown plants in the experimental field at Kyoto University (Kyoto, Japan). In addition, 30 cultivars or seedling lines with different flower colors, including yellow (‘Kidama’ and ‘Y1’), red variegation on yellow (‘Michael J’), yellow with red flush (‘Suckle Pico’), pale yellow (‘Ittosei’ and ‘16–512’), red (‘Agitato’, ‘Nekkyu’, and ‘Red Velvet’), red-white (‘OriW2’, ‘Yuino’, and ‘Matsuribayashi’), black (‘Fidalgo Blacky’, ‘Ms. Noir’, ‘Kokucho’, ‘Black Cat’, and ‘FK3’), black-white (‘Kazusa-shiranami’), deep purple (‘Super Girl’ and ‘Yukino’), purple (‘Cupid’, ‘Evelyn Rumbold’, and ‘Atom’), pink (‘Magokoro’, ‘Jun-ai’, and ‘Saffron’), ivory white (‘Gitt’s Attention’, ‘Zannsetsu’, ‘Hakuba’, and ‘Hakuyo’), were grown in the same area, and used for gene expression and

151 genotyping analyses. For *DvAKR1* hetero probe RNA gel blot analysis, *Coreopsis*
152 *grandiflora* 'Fairy Golden' (yellow), *Bidens* 'JuJu Gold' (yellow), *Cosmos sulphureus*
153 (yellow), *Chrysanthemum morifolium* 'Laub Fusha' (yellow), *Argyranthemum frutescens*
154 'Lemon Yellow' (pale yellow), *Tagetes patula* 'Harlequin' (yellow), *Gaillardia* ×
155 *grandiflora* 'Arizona Apricot' (yellow) and *Antirrhinum majus* (pale yellow) were used.
156 All plants were also grown in the experimental field or greenhouses of Kyoto University
157 (Kyoto, Japan), and ray florets or petals were collected for analysis.
158
159 *High-performance liquid chromatography (HPLC) analysis*
160 Fresh samples were homogenized with a mortar and a pestle under liquid nitrogen, then
161 1 mL extraction solution of 5 % hydrochloric acid in 50 % methanol was added. For
162 measurement of crude pigments, extraction solution of 10 % acetic acid in 50 % methanol
163 was added instead. The mixture was then centrifuged at 4°C for 15 min at 15,000 rpm,
164 and the supernatant was collected and diluted 5–50 times with the same solvent. For
165 hydrolysis, diluted solution was boiled at 95°C for 2 h, then 10–20 µL of the hydrolyzed
166 solution was injected in the HPLC apparatus. HPLC analysis was performed on a Hitachi
167 L-7100, L-7200, L-7420, and L-7500 (Hitachi Systems, Ltd., Tokyo, Japan) or with
168 HPLC Shimazu series, SCL-10AVP, SPD-M10AVP, CTO-10AVP, SIL-10ADVP, LC-
169 10ADVP, FCV-10ALVP, and DGU-14A (LCsolutions software; Shimazu Corp., Kyoto,
170 Japan). A C18 column (4.6 mm x 250 mm) (Nihon Waters K.K., Tokyo, Japan)
171 maintained at 40°C was used. The detection wavelengths were 290 nm for liquiritigenin,
172 350 nm for flavones, 380 nm for 6'-deoxychalcones and flavanols, and 520 nm for
173 anthocyanidins. Eluant preparation and HPLC analysis was performed as described by
174 Ohno et al. (2011b). Standards chemicals for quantification of hydroxylated pigments,
175 including liquiritigenin, apigenin and luteolin, kaempferol, quercetin, and cyanidin

chloride were obtained from Wako Pure Chemical Industries. Isoliquiritigenin and butein were obtained from Tokyo Chemical Industry Co. Pelargonidin was extracted by thin layer chromatography, and delphinidin chloride was purchased from Nagara Science, Gifu, Japan. For analysis of crude pigments, ray florets data of 'SH' (Ohno et al. 2021) were used as standard. Quantification of flavonoids in ray florets were performed in triplicates.

RNA-seq analysis

Total RNA in 'SH' and 'RK' were extracted from ray florets at developmental stage 2 (Fig. 2b). Two RNA samples for each cultivar were mixed equally. The mixed RNA samples were sequenced using Illumina Hiseq2000 with 101-bp paired end. A total of 54,314,990 and 42,340,142 trimmed reads were obtained in 'SH' and 'RK', respectively. Data from each of the four libraries were *de novo* assembled by Trinity (Grabherr et al. 2011), and differentially expressed genes were analyzed by RSEM-based abundance estimation. Contigs of top 1000 FPKM values were selected to draw a scatter plot diagram.

Isolation of genomic DNA and genotyping

Genomic DNA was extracted from leaves or ray florets using MagExtractor Plant Genome (Toyobo, Osaka, Japan). Transcripts or genomic fragments were cloned into pTAC-1 vector (BioDynamics Laboratory Inc., Tokyo, Japan) for sequencing. Plasmids were extracted using Quick Gene Plasmid Kit S II (Kurabo, Osaka, Japan). For inverse PCR to obtain 5′ flanking sequence, *Dra*I, *Eco*RV, *Hind*III, and *Sac*I were used. Variation in transcript sequences of *DvAKR1*, *DvCHS2*, *DvCH3H*, and genomic sequence of *DvAKR1* from 'SH,' 'IB,' and 'RK' were compared. Primers used in this analysis were designed from RNA-seq data or using previously reported sequences, and are shown in

Table S1.

In addition to ‘SH’, ‘IB’, and ‘RK’, genotyping of *DvAKR1* was also conducted in the 30 cultivars or seedling lines. Genomic DNA was extracted with MagExtractor Plant Genome (Toyobo). PCR was performed with KOD-FX Neo polymerase system (Toyobo), with the program set as follows: initial denaturation at 94°C for 2 min, followed by 30 cycles at 98°C for 10 s, 55°C for 30 s, and a final extension at 68°C for 2 min. Genotyping primers are shown in Table S2.

Gene expression analysis

Total RNA was extracted from ray florets of ‘SH’, ‘IB’, and ‘RK’ at different developmental stages, as well as in folded ray florets of the 30 cultivars or seedling lines using Sepasol RNA I Super G (Nacalai Tesque, Kyoto, Japan). Samples were purified with a High-salt precipitation solution (Takara Bio Inc., Ohtsu, Japan), then reverse transcribed (RT) with ReverTra Ace (Toyobo). The relative expression of *DvAKR1* gene was subsequently investigated using qRT-PCR. About 2 µL of 50-fold or 1 µL of 5-fold diluted RT product was used as a template for qRT-PCR. qRT-PCR was performed with SYBR Premix Ex Taq™ II (Takara Bio Inc.) or THUNDERBIRD SYBR qPCR Mix (Toyobo) according to manufacturer’s instructions in a LightCycler 480 system (Roche Diagnostics K.K., Tokyo, Japan), then run with programs set as follows: initial denaturation at 95°C for 5 min, followed by 45 cycles at 95°C for 10 s, and a final extension at 60°C for 30 s, or with an initial denaturation at 95°C for 2 min, followed by 40 cycles at 95°C for 10 s, 55°C for 5 s, and a final extension at 72°C for 20 s. Single-target product amplification was checked using a melting curve. The primers used for qRT-PCR are shown in Table S3.

For RT-PCR analysis of *DvCHS2*, *DvCH3H*, *DvAKR2*, *DvAKR3*, *DvAKR4*, and

DvAKR5 genes, total RNA was extracted from ray florets at developmental stage 2 using Sepasol RNA I Super G (Nacalai Tesque). After reverse transcription with ReverTra Ace (Toyobo), PCR was performed with Blend Taq polymerase (Toyobo). The PCR was set as follows: initial denaturation 94°C for 2 min, followed by 30–35 cycles at 94°C for 30 s, 55°C for 30 s, and a final extension at 72°C for 2 min. Primers used for RT-PCR are shown in Table S4.

For analysis of *DvAKR1* transcript sequence, RT-PCR products of ray florets at developmental stage 2 obtained with DvAKR1 Full-F and DvAKR1 Full-R primers (Table S1) were digested with *Sac*I (Takara Bio Inc). The reaction mixture consisting of 2.5 µL PCR product, 0.1 µL *Sac*I, 0.5 µL 10X L reaction buffer, and 1.9 µL distilled water was prepared and incubated overnight at 37°C.

Phylogenetic analysis

A phylogenetic tree was constructed with the open reading frames (ORFs) or amino acid sequences of different *AKR* and polyketide synthase genes using neighbor-joining method, and bootstrap consensus inferred with 1,000 replicates in MEGA 11 (https://www.megasoftware.net) (Tamura et al. 2021; Saitou and Nei 1987). The accession numbers for amino acid sequences used in the phylogenesis were as follows: 2A1: *Malus* x *domestica* NADP-dependent D-sorbitol-6-phosphate dehydrogenase (S6PDH), (P28475); 4A1: *Glycine max* CHR, (CAA39261); 4A2: *Medicago sativa* CHR, (CAA57782); 4A3: *Glycyrrhiza echinata* polyketide reductase (PKR), (BAA12084); 4A4: *Glycyrrhiza glabra* PKR, (BAA13113); 4B1: *Sesbania rostrata* CHR, (CAA11226); 4B2: *Papaver somniferum* codeinone reductase (CodR), (AAF13739); 4B3: *Papaver somniferum* CodR, (AAF13736); 4B4: *Fragaria* x *ananassa* D-galacturonate reductase (GalUR), (AAB97005); 4B5: *Zea mays* deoxymugineic acid

synthase (DAS), (BAF03164); 4C1: *Hordeum vulgare* aldose reductase (ADR), (P23901); 4C3: *Avena fatua* ADR, (Q43320); 4C5: *Digitalis purpurea* ADR, (CAC32834); 4C8: *Arabidopsis thaliana* AT2g37760, (ABH07514); 4C10: *Arabidopsis thaliana* AT2G37790, (ABH07516); 6C1: *Arabidopsis thaliana* AT1G04690, (AAA87294); DvAKR1-1, (BDE26431); DvAKR1-2, (BDE26432); DvAKR1-3, (BDE26433); DvAKR1-4, (BDE26434); DvAKR2, (BDE26435); DvAKR3, (BDE26436); DvAKR4, (BDE26437); DvAKR5, (BDE26438), *Arabidopsis thaliana* AT1G60710, (OAP19317); *Fragaria x ananassa* AKR, (AAV28174); *Glycine max* CHR4, (AIT97303); *Glycine max* CHR5, (NP_001353935); *Glycine max* CHR6, (BBC21043); *Perilla setoyensis* Alcohol Dehydrogenase (AlDehy), (AFV99150); *Rauvolfia serpentine* Perakine reductase (PR), (AAX11684); *Vitis vinifera* Galacturonic acid reductase (GalAcRed), (NP001268125); and *Zea mays* AKR2, (PWZ18047). Classification of AKRs were consistent with their clustering in the Aldo-Keto Reductase Superfamily database (https://www.med.upenn.edu/akr/).

Production of stable transgenic tobacco plants

The cDNAs of *DvAKR1-1*, *DvAKR1-2*, *DvCHS2, DvCH3H, GmCHR5, GmCHR6*, and *Am4'CGT* were first subcloned into pDONR221 vector (Invitrogen, Carlsbad, CA, USA) then transformed into a pGWB2 Gateway binary vector (Nakagawa et al. 2007). *DvAKR1-1*, *DvAKR1-2,* and *DvCH3H* were cloned from the ray florets of 'SH', while *DvCHS2* was cloned from ray florets of dahlia 'Yuino' (Ohno et al. 2018a). Similarly, *GmCHR5* and *GmCHR6* were cloned from bean sprouts of *Glycine max*, while *Am4'CGT* was cloned from the yellow flower of *Antirrhinum majus*. Transgenic tobacco plants were obtained by standard *Agrobacterium tumefaciens* EHA105 strain leaf disk transformation method as described by Horsch et al. (1989) using *Nicotiana tabacum*, which is able to

276 synthesize flavonoids (anthocyanins and flavonols) in flowers. The generated transgenic
277 T_0 plants harboring a copy of the transgenes were selected by genomic PCR, qPCR, or
278 DNA gel blot analysis. For genotyping, genomic DNA was extracted using SDS method.
279 qPCR was performed with THUNDERBIRD SYBR qPCR Mix (Toyobo) using the
280 following program: initial denaturation at 95°C for 2 min, followed by 40 cycles at 95°C
281 for 10 s, 55°C for 5 s, and a final extension at 72°C for 20 s. Single-target product
282 amplification was checked using a melting curve. *NtActin* was used as the internal
283 standard, and transcript abundance was calculated by relative quantification using a
284 standard curve. DNA gel blot analysis was conducted according to Ohno et al. (2018a).
285 Transgenic T_0 plants were self-crossed to obtain T_1 generation plants harboring
286 homozygous transgenes, which were screened by PCR and qPCR. Multiple transgenic
287 overexpression lines were obtained by crossing T_0 plants or T_1 plants. pGWB2-*GUS*
288 overexpression lines were used as a mock treatment. Primers used for gateway cloning,
289 genotyping of transgenes, and expression analysis of transgenes are shown in Table S5,
290 Table S6, and Table S7, respectively. Primers for *Am4'CGT* were designed according to
291 Hoshino et al. (2019).
292
293 *Agroinfiltration assay*
294 Transient expression of 6'- deoxychalcone biosynthetic genes was examined in tobacco
295 to verify their functional activities. cDNA of *DvAKR1-1, DvAKR1-2*, *DvCHS2, DvCH3H,*
296 *GmCHR5, GmCHR6, Am4'CGT*, and *CaMYBA* were subcloned in pDONR221 vector
297 (Invitrogen) then recombined in pGWB2 Gateway binary vector (Nakagawa et al. 2007).
298 *CaMYBA* was cloned from the purple flowers of *C. annuum* 'Peruvian Purple' (Ohno et
299 al. 2020). All constructs were transformed into *A. tumefaciens* EHA105 strain. The *BETA-*
300 *GLUCURONIDASE* (β *-GUS*) gene was used in control assays.

Transient over-expression in benthamiana tobacco plants was performed according to the methods of Polturak et al. (2016). Transgenic *A. tumefaciens* were infiltrated directly or co-infiltrated in 3–4 weeks old seedling leaves. For co-infiltration, each *Agrobacterium* suspension with 0.2–1.0 cell density at OD_{600} were mixed in equal ratios before infiltration. For co-expression of *DvAKR1/GmCHR5, DvCH3H, Am4'CGT*, and *CaMYBA*, the tomato bushy stunt virus (TBSV) p19 silencing suppressor expressed in the pDGB3alpha2_35S:P19:Tnos (GB1203) vector (addgene) was also mixed before infiltration. Leaves used for pigment extraction and RT-PCR were sampled between the fifth and the seventh day post infiltration. Samples were obtained in three biological replicates consisting of at least three different leaves for each experiment. Primers used for RT-PCR are shown in Table S7.

The *GUS* expression was confirmed by GUS staining. Briefly, *GUS* infiltrated leaves were treated with GUS buffer (0.5 mM potassium ferricyanide, 0.5 mM potassium ferrocyanide, 0.3% TritonX-100, 5.0% methanol, 50 mM/pH 7.0 NaH_2PO_4 suspended in distilled water). The GUS staining buffer (1.0 mM X-glucuronide, 0.5 mM potassium ferricyanide, 0.5 mM potassium ferrocyanide, 0.3% tritonX-100, 5.0% methanol, 50 mM/pH 7.0 NaH_2PO_4 suspended in distilled water) was then used to infiltrate leaves for overnight staining at 37℃. Finally, the leaves were decolorized by washing with 70% ethanol.

Hetero-probe RNA gel blot analysis

Total RNA was extracted from flowers using Sepasol RNA I Super G (Nacalai Tesque) and purified by precipitation with a High-salt solution (Takara Bio Inc.). Approximately 5 μg of total RNA was used for gel blot analysis. Full length CDS of *DvAKR1* was labeled with the digoxigenin (DIG) RNA Labeling Kit (Sigma-Aldrich, St. Louis, USA). The

probe was hybridized to the membrane overnight at 50ºC. Detection of protein bands were visualized with CDP-Star chemiluminescence substrate (GE Healthcare Japan, Tokyo, Japan) followed by imaging with a LAS-3000 Mini (Fujifilm, Tokyo, Japan).

Statistical analysis

Data were expressed in each treatment as mean ± standard error of three biological replicates. Statistical significance was assessed using Tukey's honest significant difference test in excel for windows version 5.0, and a probability value less than 0.05 ($P < 0.05$) was considered significant.

Results

Biosynthesis of 6'-deoxychalcone is lost in the ray florets of 'RK'

In this study, a red-white bicolor dahlia cultivar 'SH' and its two-lateral mutants, 'IB' and 'RK' were used (Fig. 2a). Accumulation of anthocyanins, flavones, and 6'-deoxychalcones, such as butein and isoliquiritigenin was detected in its red ray florets of 'SH'. Similar pigments were also accumulated in 'IB', which show deeper red color than 'SH', with higher anthocyanidins and flavones, but lower 6'-deoxychalcones contents (Fig. 2c-e). 'RK' which produces purple flowers accumulated anthocyanins and flavones but not 6'-deoxychalcones (Fig. 2c–e). A two-step mutation from mutation of dahlia cultivars 'SH' to 'IB' and 'IB' to 'RK' exhibited a gradual decrease in the content of 6'-deoxychalcones (Fig. 2e). Since butein is presumably biosynthesized from

isoliquiritigenin, we speculated that mutation in the genes involved in isoliquiritigenin biosynthesis has occurred in ‘IB’ and ‘RK’.

The loss of 6'-deoxychalcone biosynthesis in ‘RK’ is unlikely to be associated with CHS, CH3H, and AKR 4B sub-family genes

Previous studies suggest that *DvCHS2* is involved in the anthocyanin, flavone, and 6'-deoxychalcone biosynthesis in dahlia (Ohno et al. 2011b; 2018a). Since *CH3H* gene had previously not been isolated in dahlia, our study identified a *DvCH3H* gene sharing 90% homology with *C. sulphureus CsCH3H* gene, and containing a substrate recognition site (SRS1) and XSAGGXX domain (Fig. S1) (Schlangen et al. 2010b). An allelic variant of *F3'H* was previously shown to have a *CH3H* activity, and valine at position 425 was identified to be crucial for chalcone substrate acceptance (Schlangen et al. 2010a). Interestingly, *DvCH3H* identified in this study also harbored valine at position 425 (Fig. S1). In addition, gene expression analysis showed no difference in the profiles of *DvCHS2* and *DvCH3H* in ‘SH’ and ‘RK’ (Fig. S2). We analyzed *DvCHS2* and *DvCH3H* cDNA sequences, however no difference was found between ‘SH’ and ‘RK’. Thus, *DvCHS2* and *DvCH3H* do not explain the difference of 6'-deoxychalcone biosynthesis in these two cultivars. Therefore, *CHR* gene or its equivalent is potentially the causal gene associated with the loss of 6'-deoxychalcone biosynthesis in ‘RK’.

*CHR*s are key genes involved in the isoflavone biosynthetic pathway in legumes, and are members of the AKR 4A sub-family, which uniquely contains legume *CHRs* but not aldo-keto reductase genes (Fig. 3a). Notably, no AKR 4A sub-family related genes in the ray florets were identified by RNA-seq in this study. However, four AKR 4B sub-family genes, *DvAKR2-DvAKR5*, which are phylogenetically adjacent to *AKR* genes were expressed in the ray florets (Fig. 3a). Gene expression analysis of these four genes

revealed no difference in their profiles in ‘SH,’ ‘IB,’ and ‘RK’ (Fig. S2). Together, these results indicated that novel gene that is completely different from legume *CHRs* is involved in the isoliquiritigenin biosynthesis in dahlia.

A candidate c25599_g2_i1 (DvAKR1) gene is associated with isoliquiritigenin biosynthesis in dahlia

To explore candidate genes of isoliquiritigenin biosynthesis in dahlia, a comparative RNA-seq analysis of developmental stage 2 ray florets of ‘SH’ and ‘RK’ was conducted, and a total 106,692 contigs were obtained. Of these, 74,045 contigs were translated into amino acid sequences by *de novo* assembly. Among top 1,000 contigs based on mean FPKM values in ‘SH’ and ‘RK’, a candidate contig, *c25599_g2_i1*, was differentially expressed, with a 9.5-fold higher profile in ‘SH’ than in ‘RK’ (Fig. 4a; Table S8). In contrast, the differences in FPKM value of other flavonoid-related genes, including *DvCHS1*, *DvCHS2*, and *DvCH3H* were less than 2-fold higher or lower between ‘SH’ and ‘RK’ (Table S8). The candidate contig shared high homology with AKR, thus it was designated *DvAKR1*. Subsequently, analysis of *DvAKR1* expression levels at different developmental stages in ‘SH’, ‘IB’ and ‘RK’ showed that its profiles were significantly highly expressed in ‘SH’ at stage 2 and stage 3 (Fig. 4b). In addition, the *DvAKR1* expression analysis in the tested 30 cultivars or seedling lines with various flower colors revealed that its profiles were relatively higher in 6'-deoxychalcone producing cultivars, but lower or almost undetectable in non 6'-deoxychalcone producing cultivars (Fig. 4c). However, higher expression of *DvAKR1* in a non-6'-deoxychalcone accumulating sample was only observed in the ‘Super Girl’ cultivar. These results further confirmed that *DvAKR1* is a candidate gene associated with isoliquiritigenin biosynthesis in dahlia.

The *DvAKR1* transcript contained a full length ORF of 1065 bp, encoding 354

amino acid residues. BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) revealed that *DvAKR1* is a member of the AKR superfamily, which is distant to the AKR 4A sub-family that contains legume *CHRs* (Fig. 3a). Putative amino acid sequence of *DvAKR1* shared relatively high homology of 69%, 55%, and 64% with AKR 13 sub-family orthologs, such as AT1G60710 from *Arabidopsis*, perakine reductase from *R. serpentine*, and alcohol dehydrogenase from *Perilla frutescens*, respectively (Fig. 3b). In addition, DvAKR1 contained conserved catalytic tetrad, Asp^{57}, Tyr ^{62}, Lys^{88}, and His^{130}, and several typical AKR cofactor binding residues, including Ser^{160}, Gln^{180}, and Tyr^{208} (Fig. 3a) (Sun et al. 2012).

Genomic sequence of *DvAKR1* contains five exons and four introns. The genomic structure of *DvAKR1* was similar to *AT1G60710*, which also consist of five exons and four introns, but different from *GmCHR1* which consist of three exons and two introns (Fig. 5a). Dahlia is an octoploid, thus, multiple *DvAKR1* alleles were predicted, consequently, four different *DvAKR1* sequences designated *DvAKR1-1* to *DvAKR1-4*, were identified in 'SH' and 'RK' based on the SNPs and indels, with sequences of *DvAKR1-3* exhibiting a deletion in the last exon, and encoding a truncated protein (Fig. 5b). A high homology of 88–98 % in the putative *DvAKR1* amino acid sequence was observed (Fig. 5c). An inverse PCR to isolate the *DvAKR1* 5' flanking sequence in 'SH' and 'RK', identified about 320 bp region upstream of the start codon for *DvAKR1-1* and *DvAKR1-2*, and about 700 bp region for *DvAKR1-3* and *DvAKR1-4*. Notably, no differences were observed in these identified sequences between 'SH' and 'RK'.

To analyze the predominantly expressed *DvAKR1* gene in 'SH,' 'IB,' and 'RK', a combined RT-PCR with *Sac*I digestion was performed using ray floret samples at stage 2. The *Sac*I restriction of *DvAKR1-1* and *DvAKR1-2* was in the middle of exon three, while *Sac*I restriction site was absent in *DvAKR1-3* and *DvAKR1-4* (Fig. 5a). Though a

425 more intense undigested band was detected in ‘RK’ than ‘SH’, strong *Sac*I digested bands
426 with faint undigested bands were observed in all the three cultivars, indicating *DvAKR1-*
427 *1* and *DvAKR1-2* are much highly expressed than *DvAKR1-3* and *DvAKR1-4* in these
428 dahlia cultivars (Fig. 5d).

429 The genetic background of *DvAKR1* was analyzed in 30 cultivars or seedling
430 lines with genomic PCR, using P1/P3 or P2/P3 primer pairs to detect *DvAKR1-1* and
431 *DvAKR1-2* or *DvAKR1-3* and *DvAKR1-4*, respectively (Fig. 5a). As a result, P1/P3
432 amplified bands were detected in ‘SH’, ‘IB’, and ‘RK’, and in cultivars accumulating 6'-
433 deoxychalcones. In contrast, P1/P3 amplified bands were not detected in all cultivars
434 lacking 6'-deoxychalcones accumulation, except ‘Super Girl’ and ‘Magokoro’ (Fig. 5e).
435 This result was correlated with *DvAKR1* expression data, which indicated that with
436 exception of ‘RK’, ‘Super Girl’ and ‘Magokoro’, cultivars harboring *DvAKR1-*
437 *1*/*DvAKR1-2* had higher expression of *DvAKR1* and 6'-deoxychalcone accumulation.
438 From these results, we speculated that *DvAKR1-1* and *DvAKR1-2* were candidate genes
439 involved in 6'-deoxychalcone synthesis in dahlia.

440

441 *Isoliquiritigenin accumulation is not induced by over-expression of DvAKR1-1, DvAKR1-*
442 *2, GmCHR5, or GmCHR6 in transgenic tobacco*

443 For functional validation of *DvAKR1*, transgenic tobacco plants overexpressing *DvAKR1-*
444 *1* or *DvAKR1-2* were generated. Several over-expression T_1 lines were obtained, however,
445 flower colors appeared to be similar to *GUS* overexpressing plants (Fig. 6a). Though we
446 detected introduced gene expressions (Fig. 6b), however only flavonol, and
447 isoliquiritigenin nor butein were not detected in the flowers (Fig. 6c). Similarly, no
448 isoliquiritigenin or butein were detected in the flowers of positive over-expression lines
449 produced using functional soybean *CHRs*, including *GmCHR5* or *GmCHR6* (Fig. 6c)

(Mameda et al. 2018). These results suggested that single over-expression of *CHR* is not sufficient to induce isoliquiritigenin or butein accumulation in tobacco. In addition, transgenic lines overexpressing *DvCHS2* or *DvCH3H* were crossed with those overexpressing *DvAKR1-1*, *DvAKR1-2* or *GmCHR5* to express multiple butein biosynthesis pathway genes. However, only kaempferol was detected and neither isoliquiritigenin nor butein was detected in the resulting flowers (Fig. 6c). Interestingly, liquiritigenin was previously detected in transgenic tobacco flowers overexpressing *Pueraria montana* chalcone reductase (Joung et al. 2003). However, no liquiritigenin was detected in all tested transgenic lines in this study (Fig. S3).

Transient co-overexpression of DvAKR1/GmCHR, Am4'CGT and CaMYBA sufficiently induce isoliquiritigenin accumulation in benthamiana tobacco leaves

Generally, flavonoids are accumulated *in vivo* as glycosides, and dahlia accumulates isoliquiritigenin and butein as 4′-glucoside or its derivatives (Harborne et al. 1990; Ohno et al. 2021). We assumed that tobacco plants are unable to accumulate isoliquiritigenin without glycosylation *in vivo* due to lack of suitable glucosyltransferase, thus, we attempted to co-overexpress either *DvAKR1-1*, *DvAKR1-2*, *GmCHR5*, or *GmCHR6* with *Am4'CGT*, which was selected because *chalcone glucosyltransferase* gene is yet to be isolated in dahlia, and the site of its glycosylation is predicted to be in the same position as those of isoliquiritigenin 4′-glucoside and butein 4′-glucoside. Co-overexpression of either *DvAKR1-1*, *DvAKR1-2*, *GmCHR5*, or *GmCHR6* with *Am4'CGT* was performed by transient agroinfiltration in *N. benthamiana* leaves. As a result, both isoliquiritigenin and flavonols were not detected in the infiltrated leaves, indicating that flavonoid biosynthesis activity is very low in *N. benthamiana* leaves (Fig. 7c). To rescue flavonoid accumulation in tobacco leaves, co-expressions of either *DvAKR1-1*, *DvAKR1-2*, *GmCHR5*, or

GmCHR6 and *Am4'CGT*, were double co-overexpressed with a MYB TF, *CaMYBA*, that positively regulates anthocyanin biosynthesis in pepper (Borovsky et al. 2004; Ohno et al. 2020). Infiltrated leaves turned reddish (Fig. 7a) and introduced gene expressions were detected (Fig. 7b). As a result, isoliquiritigenin in the leaves was detected in the co-infiltrated assay involving *DvAKR1-1* or *DvAKR1-2* with both *Am4'CGT* and *CaMYBA* at a frequency of 3/8 and 2/8 for *DvAKR1-1* and *DvAKR1-2*, respectively (Fig. 7c and 7d). In addition, isoliquiritigenin in the leaves was detected in co-infiltrated assay of *GmCHR5* or *GmCHR6* with both *Am4'CGT* and *CaMYBA* at a frequency of 3/8 and 2/8 for *GmCHR5* and *GmCHR6*, respectively, while not detection was observed from *Am4'CGT* and *CaMYBA* co-infiltrated leaves (Fig. 7c). Moreover, liquiritigenin and delphinidin were also detected in either *DvAKR1-1*, *DvAKR1-2*, *GmCHR5* or *GmCHR6*, *Am4'CGT* and *CaMYBA* co-infiltrated leaves (Fig. S3 and S4). We also analyzed crude leaf extracts of either *DvAKR1-1*, *DvAKR1-2*, *GmCHR5*, or *GmCHR6*, co-overexpressed with both *Am4'CGT* and *CaMYBA*. Comparison of our data with those of previously reported 'SH' ray florets extract profiles (Ohno et al. 2021) revealed the detection of peaks corresponding to isoliquiritigenin 4'-*O*-glucoside and isoliquiritigenin 4'-*O*-[6-*O*-(malonyl)-glucoside] (Fig. S5). Interestingly, co-overexpression of *DvAKR1-1* or *DvAKR1-2* with either *CaMYBA* and *GmCHR5* or *GmCHR6* and *CaMYBA* revealed no induction of isoliquiritigenin accumulation (Fig. 7c). In addition, a cross between transgenic tobacco lines overexpressing *Am4'CGT* and either over-expression lines of *DvAKR1-1*, *DvAKR1-2*, *GmCHR5*, or *GmCHR6* to identify genes showed no isoliquiritigenin detection in flowers (Fig. 6c), which indicated that unknown genes under the regulation of *CaMYBA* are essential for isoliquiritigenin accumulation in tobacco. Overall, these results suggested that *DvAKR1* is involved in isoliquiritigenin biosynthesis, with chalcone glycosylation and other factors under the regulation of *CaMYBA* as crucial

processes in the tobacco leaf isoliquiritigenin accumulation.

Finally, *DvAKR1/GmCHR5, Am4'CGT,* and *CaMYBA* were co-overexpressed with *DvCH3H* in combination with p19 suppressor protein of TBSV to induce butein accumulation successfully resulted in the detection of butein in co-infiltrated leaves (Fig. 8).

Asteraceae species accumulating 6'-deoxychalcone or 4-deoxyaurone show homologous DvAKR1 gene expression

Yellow flower color of Asteraceae species is derived from carotenoids or 6'-deoxychalcones. In snapdragon, 4-hydroxyaurones are directly synthesized from naringenin chalcone (Ono et al. 2006), however in Asteraceae species, polyphenol oxidase catalyzes the synthesis of 4-deoxyaurone from butein (Molitor et al. 2015, 2016). To analyze the involvement of *DvAKR1* orthologs in the accumulation of 6'-deoxychalcone and 4-deoxyaurone in ray florets of other Asteraceae species, we performed hetero-probe RNA gel blot analysis using the full length CDS of *DvAKR1*. Butein or aurone biosynthesis plant group, including dahlia, coreopsis, bidens, and *C. sulphureus* were selected, while chrysanthemum, *A. frutescens*, marigold (*Tagetes patula*), and *Gaillardia* × *grandiflora* species were selected as non-butein biosynthesis group. As a result, intense bands were detected in 6'-deoxychalcone and 4-deoxyaurone producing plant group, while no bands were detected in the non-producing plant group as well as in snapdragon (Fig. 9). Therefore, a clear correlation in *DvAKR1* homologous gene expression and 6'-deoxychalcone and 4-deoxyaurone accumulation was observed, suggesting that its homologous expression might be crucial for 6'-deoxychalcone biosynthesis in dahlia and other 6'-deoxychalcone producing Asteraceae species.

Discussion

DvAKR1 is a potential determinant of 6′-deoxychalcone biosynthesis in dahlia

Butein is chemically known as 2′,3,4,4′-tetrahydroxychalcone, and the name is derived from the genus *Butea*. Butein is found in limited species, such as dahlia and coreopsis from Asteraceae, *Searsia* from Anacardiaceae, and in Fabaceae (Semwal et al. 2015). As a flower color determinant, 6'-deoxychalcones are among the few important flavonoid-related compounds exhibiting the bright yellow color. Carotenoids or betaxanthins confers the yellow flower color in many species, thus, plant species lacking flavonoid-related compounds rarely have yellow flower color. To develop yellow flower through transgenic approach in such species, inducing the biosynthesis of 6'-deoxychalcones rather than carotenoids or betaxanthins is more effective and convenient method. Therefore, revealing and understanding the 6'-deoxychalcone biosynthetic pathway is crucial. Butein is synthesized via a two-step process that begins with isoliquiritigenin biosynthesis, which is catalyzed by CHS and AKR followed by the conversion of isoliquiritigenin to butein by CH3H (Fig. 1). In dahlia ray florets, butein is co-synthesized with isoliquiritigenin (Fig. 2e; Ohno et al. 2011b, 2013a), suggesting that isoliquiritigenin biosynthesis is the determinant step. Given that CHS is a common enzyme associated with other flavonoids, such as anthocyanins and flavones, it is likely that AKR is the determinant enzyme for 6'-deoxychalcone biosynthesis in dahlia.

Dahlias are octoploid and highly heterologous plants, and their mutant cultivars are ideal materials for comparative analysis. Thus, lateral mutant cultivars ‘SH’, ‘IB’ and ‘RK’ were used in this study. Flower color of ‘SH’ is a red-white bicolor, which

accumulate 6'-deoxychalcones, anthocyanins, and flavones. Based on the decrease in 6'-deoxychalcone and the increase of anthocyanin and flavone content, the colored flower parts of 'IB' and 'RK' turns to dark red and purple, respectively (Fig. 2a). A comparative RNA-seq analysis identified a candidate *DvAKR1* gene, and its expression was correlated with 6'-deoxychalcone accumulation in 'SH', 'IB', 'RK', and other cultivars (Fig. 4b and 4c). Four *DvAKR1* sequences, including *DvAKR1-1*/*DvAKR1-2* were identified in all 6'-deoxychalcone producing cultivars, while all cultivars lacking 6'-deoxychalcones except for 'Super Girl' and 'Magokoro', did not harbor *DvAKR1-1*/*DvAKR1-2* (Fig. 5e). These results indicate that *DvAKR1* might be crucial for 6'-deoxychalcone biosynthesis in most dahlia cultivars. Since 'Super Girl' harbored *DvAKR1-1*/*DvAKR1-2*, and showed higher transcript level of *DvAKR1* (Fig. 4c), it was subjected to sequencing. Despite containing several SNPs, *DvAKR1* is seemingly a functional transcript, suggesting that other genes rather than *DvAKR1* are involved in the loss of 6'-deoxychalcone in 'Super Girl'. A previous genetic analysis proposed that the Y factor, which is dominant to y or I (ivory white), acts like tetrasomic, and determines the yellow color of octoploid dahlia (Lawrence 1931). Our results suggest that *DvAKR1* is a strong candidate for Y factor, which is still yet to be isolated.

Lower expression level of *DvAKR1* was observed in 'RK' than 'SH' (Fig. 4b, Fig. 5d). However, no mutation in *DvAKR1*, including in the promoter region was detected, indicating that *DvAKR1* is unlikely to be associated with the flower color change in these lateral mutant cultivars. RNA-seq analysis also showed that the expression level of *DvCHS2* was also reduced to about half (Table S8), which suggested that the loss of 6'-deoxychalcone biosynthetic ability in 'RK' caused a negative feedback regulation leading to suppressed *DvAKR1* and *DvCHS2* gene expression.

DvAKR1 and legume CHRs demonstrate independent evolutionary origins

The AKR 4A sub-family exclusively contain *CHRs* members from legume plants (Jez et al. 1997; Bomati et al. 2005). Consistently, no AKR 4A sub-family genes in dahlia ray florets were identified from our RNA-seq data, confirming that the sub-family is specific to legume isoflavonoid synthesis pathway. In contrast, four AKR 4B sub-family genes, *DvAKR2-DvAKR5*, were identified, but their expression levels did not correlate with butein accumulation (Fig. S2), indicating that they were not involved in isoliquiritigenin biosynthesis in dahlia. Similar observation showing by that the expression levels of CHR-like genes, such as CHRl1-CHRl3 could not explain butein biosynthesis has previously been reported (Walliser et al. 2021).

DvAKR1 belongs to AKR13 sub-family, and is phylogenetically distant from legume CHRs (Fig. 3a), sharing only 19% protein identity with putative GmCHR1 amino acid sequence. However, DvAKR1 is phylogenetically located adjacent to *Arabidopsis* AT1G60710, *Zea mays* AKR2, *Rauvolfia serpentine* perakine reductase and *Perilla setoyensis* PsAKR. AKR superfamily enzymes are NAD(P)(H) binding oxidoreductases that metabolize a wide array of chemical substrates (Jez et al. 2001). PsAKR catalyze the conversion of either geraniol to citral and nerol, or perilla alcohol into perillaldehyde (Sato-Masumoto and Ito 2014). The NADPH-dependent step in a side-branch of the 10-step ajmaline alkaloid biosynthetic pathway is catalyzed by perakine reductase (Sun et al. 2008). *DvAKR1* homologous gene expression was detected in the petals of Asteraceae species accumulating 6'-deoxychalcone and/or 4-deoxyaurone, but not in those lacking 6'-deoxychalcone and/or 4-deoxyaurone, which demonstrated that *DvAKR1* homologs are involved in isoliquiritigenin biosynthesis in these species (Fig. 9). Notably, these species belong to the Coreopsideae tribe, indicating that their 6'-deoxychalcone and/or 4-deoxyaurone biosynthesis pathways share *DvAKR1* gene homologs. Ecological

importance of 6'-deoxychalcones and 4-deoxyaurones are yet to be reported, however, it was indicated that isoliquiritigenin biosynthesis of legumes and Coreopsideae tribe species have independent evolutionary origins.

Co-overexpression of DvAKR1/GmCHR with Am4'CGT and CaMYBA genes induces isoliquiritigenin accumulation in tobacco

Co-overexpression of either *DvAKR1-1*, *DvAKR1-2*, *GmCHR5,* or *GmCHR6*, with both *Am4'CGT* and *CaMYBA* successfully induced isoliquiritigenin accumulation in tobacco leaves, while no induction was detected in *DvAKR1-1*, *DvAKR1-2*, *GmCHR5* or *GmCHR6* leaves co-overexpressed with only *CaMYBA* (Fig. 7c), suggesting that glycosylation process is crucial for *in vivo* accumulation of isoliquiritigenin. Am4′CGT was initially isolated as a UDP-glucose: chalcone 4'-O-glucosyltransferase for aurone formation in snapdragon (Ono et al. 2006), and *Am4'CGT* silencing represses aurone production (Bradley et al. 2017). Enzymatic assay revealed that 2′, 4′, 6′, 4-tetrahydoroxychalcone was more efficiently glycosylated than 2′, 4′, 6′, 3′ 4-pentahydoroxychalcone by Am4′CGT (Ono et al. 2006). Interestingly, our results show that Am4′CGT can potentially glycosylate 2′,4,4′-trihydroxychalcone (isoliquiritigenin) as well. To our knowledge, no previous reports have successfully shown the accumulation of isoliquiritigenin or butein by ectopic expression of legume *CHRs* in transgenic tobacco. However, two studies reported successful accumulation of isoliquiritigenin in transgenic petunia flowers overexpressing *CHR*. One is that introduction of *Medicago sativa*, *MsCHR7* cDNA induced the accumulation of butein 4-glucoside and butein 3-glucoside in flowers, isoliquiritigenin in pollen, and trace amount of chalcones in petunia leaves (Davies et al. 1998). The other is that the introduction of *PKR1* from *Lotus japonicus* induced the accumulation of isoliquiritigenin in trace levels

in transgenic petunia flowers (Shimada et al. 2006). These observations demonstrate that unlike in tobacco, single over-expression of *CHR* is sufficient to accumulate isoliquiritigenin in petunia. This difference could be due to endogenous genes in petunia, such as *glucosyltransferase*, which result in the accumulation of 6'-deoxychalcones. Accumulation of liquiritigenin in transgenic tobacco expressing *pl-chr* (*CHR*) from *P. montana*, and the isomerization capacity of isoliquiritigenin by tobacco chalcone isomerase to legume-like liquiritigenin has previously been reported (Joung et al. 2003). However, we could not detect liquiritigenin in transgenic tobacco flower overexpressing *DvAKR1-1* or *GmCHR5*. Interestingly, transient co-overexpression of *DvAKR1-1*/*GmCHR5* with both *Am4'CGT* and *CaMYBA* induced liquiritigenin accumulation in transgenic tobacco leaves (Fig. S3).

Our study also highlights the crucial role of *CaMYBA* ectopic expression in isoliquiritigenin accumulation. Since no flavonoid accumulation was detected in plants co-overexpressing either *DvAKR1-1*, *DvAKR1-2*, *GmCHR5*, or *GmCHR6* with *Am4'CGT*, we hypothesized that benthamiana leaves had weak flavonoid biosynthesis activity (Fig. 7c). However, transgenic *N. tabacum* plants co-overexpressing either *DvAKR1-1*, *DvAKR1-2*, *GmCHR5*, or *GmCHR6* with *Am4'CGT* accumulated only kaempferol in the flowers but not isoliquiritigenin (Fig. 6c), which indicated that unknown factors other than flavonoid biosynthesis activity was responsible for the lack of isoliquiritigenin accumulation. Interestingly, gene(s) regulated by *CaMYBA* are essential for isoliquiritigenin accumulation. *CaMYBA* belongs to the AN2 clade of MYB TFs, and virus-induced gene silencing of *CaMYBA* demonstrated its role in the regulation of *CaCHS*, *CaCHI*, *CaF3H*, *CaF3'5'H*, *CaDFR*, *CaANS*, *CaUFGT*, *CaANP*, and *CaGST* in pepper (Zhang et al. 2015). Tobacco genome encodes a floral tissue specific MYB transcription factor *NtAN2* (Pattanaik et al. 2010). Both *CaMYBA* and *NtAN2* belongs to

650 the AN2 sub-group (Borovsky et al. 2004; Pattanaik et al. 2010), however, tobacco plants
651 primarily lack delphinidin, suggesting that *CaMYBA* and *NtAN2* regulate different sets
652 of structural genes. Chen et al. (2019) reported that over-expression of *MaMYBA* from
653 grape hyacinth (*Muscari armeniacum*) in tobacco plants induced delphinidin
654 accumulation in the leaves, thus, demonstrating the differential regulation of AN2 sub-
655 group MYB TFs. In this study, ectopic expression of *CaMYBA* induced the expression of
656 one or more gene, sufficiently resulting in the accumulation of isoliquiritigenin. Further
657 studies to identify the *CaMYBA* target gene for isoliquiritigenin accumulation in tobacco
658 are needed to facilitate future molecular breeding of yellow flowers through butein
659 biosynthesis.

660 In summary, this study identified a novel *AKR* gene belonging to the AKR 13
661 sub-family, and with phylogenetically distinct features from those of known legume
662 *CHRs*. In addition, co-overexpression of *DvAKR1*/*GmCHR*, *Am4'CGT*, and *CaMYBA*
663 sufficiently induced isoliquiritigenin accumulation, while butein accumulation in tobacco
664 was induced by co-overexpression of either *DvAKR1*/*GmCHR*, *DvCH3H,* with both
665 *Am4'CGT* and *CaMYBA*. Overall, these results demonstrate the essential role of
666 *Am4'CGT* and *CaMYBA* in the metabolic engineering of 6'-deoxychalcone, 4-
667 deoxyaurone, and isoflavone biosynthetic pathways.

668

669

670

671 **Supplementary data**

672 **Table S1** Primers used for the isolation of *DvAKR1*, *DvCHS2*, and *DvCH3H* genes

673 **Table S2** Primers used for *DvAKR1* genotyping

674 **Table S3** Primers used for real-time RT-PCR

675 **Table S4** Primers used for RT-PCR

676 **Table S5** Primers used for gateway cloning

677 **Table S6** Primers used for genotyping of transgenic tobacco plants

678 **Table S7** Primers used for validation of transgenic expression

679 **Table S8** FPKM values of flavonoid biosynthesis related genes in the ray florets of 'SH'
680 and 'RK' at developmental stage 2

681 **Fig. S1** Comparison of putative amino acid sequence of *CH3H* and *F3'H* genes

682 **Fig. S2** RT-PCR analysis of *DvCHS2*, *DvCH3H*, *DvAKR2*, *DvAKR3*, *DvAKR4,* and
683 *DvAKR5* in the ray florets of 'SH', 'IB', and 'RK' developmental stage 2

684 **Fig. S3** Detection of liquiritigenin in the transgenic tobacco flower co-overexpressing
685 *DvAKR1*/*GmCHR5*/*GmCHR6* with *Am4'CGT*, and in benthamiana leaves co-
686 overexpressing *DvAKR1/GmCHR5*, *Am4'CGT* with *CaMYBA*

687 **Fig. S4** Detection of anthocyanins in benthamiana leaves co-infiltrated with either
688 *DvAKR1/GmCHR5/GmCHR6*, or both *Am4'CGT* and *CaMYBA*

689 **Fig. S5** HPLC analysis of crude extract of *DvAKR1/GmCHR5*, *Am4'CGT,* and *CaMYBA*
690 co-infiltrated benthamiana leaves

691

692

693

694 **Conflict of interests**

695 The authors declare no conflict of interests.

696

697

698

699 **Data availability statement**

700 The data that support the findings of this study are openly available in the Genbank and
701 Sequence Read Archive under the accession number PRJDB12893
702 (https://www.ncbi.nlm.nih.gov/bioproject/PRJDB12893). Accession numbers: *DvAKR1-*
703 *1* cDNA (LC671883), *DvAKR1-1* genomic sequence (LC671879), *DvAKR1-2* cDNA
704 (LC671884), *DvAKR1-2* genomic sequence (LC671880), *DvAKR1-3* cDNA (LC671885),
705 *DvAKR1-3* genomic sequence (LC671881), *DvAKR1-4* cDNA (LC671886), *DvAKR1-4*
706 genomic sequence (LC671882), *DvAKR2* (LC671887), *DvAKR3* (LC671888), *DvAKR4*
707 (LC671889), *DvAKR5* (LC671890) and *DvCH3H* (LC671891). Preliminary RNA-seq:
708 'Shukuhai' ray floret at stage 2 (DRR337625); 'Rinka' ray floret at stage 2 (DRR337626).
709
710
711
712 **References**

713 Ayabe SI, Udagawa A, Furuya T (1988) NAD(P)H-dependent 6'-deoxychalcone
714 synthase activity in *Glycyrrhiza echinata* cells induced by yeast extract. Arch
715 Biochem Biophys 261:458–462. https://doi.org/10.1016/0003-9861(88)90362-1
716 Bomati E, Austin M, Bowman M, Dixon R, Noel J (2005) Structural elucidation of
717 chalcone reductase and implications for deoxychalcone biosynthesis. J Biol Chem
718 280:30496–30503. https://doi.org/10.1074/jbc.M502239200
719 Borovsky Y, Oren-Shamir M, Ovadia R, Jong WD, Paran I (2004) The *A* locus that
720 controls anthocyanin accumulation in pepper encodes a *MYB* transcription factor
721 homologous to *Anthocyanin2* of *Petunia*. Theor Appl Genet 109:23–29.
722 https://doi.org/10.1007/s00122-004-1625-9
723 Bradley D, Xu P, Mohorianu II, Whibley A, Field D, Tavares H, Couchman M, Copsey
724 L, Carpenter R, Li M, Li Q, Xue Y, Dalmay T, Coen E (2017) Evolution of flower

color pattern through selection on regulatory small RNAs. Science 358:925–928. https://doi.org/10.1126/science.aao3526

Chen K, Du L, Liu H, Liu Y (2019) A novel R2R3-MYB from grape hyacinth, MaMybA, which is different from MaAN2, confers intense and magenta anthocyanin pigmentation in tobacco. BMC Plant Biol 19:390. https://doi.org/10.1186/s12870-019-1999-0

Davies KM, Bloor SJ, Spiller GB, Deroles SC (1998) Production of yellow colour in flowers: redirection of flavonoid biosynthesis in *Petunia*. Plant J 13:259–266. https://doi.org/10.1046/j.1365-313X.1998.00029.x

Deguchi A, Ohno S, Hosokawa M, Tatsuzawa F, Doi M (2013) Endogenous post-transcriptional gene silencing of flavone synthase resulting in high accumulation of anthocyanins in black dahlia cultivars. Planta 237:1325–1335. https://doi.org/10.1007/s00425-013-1848-6

Gatt M, Ding H, Hammett K, Murray B (1998) Polyploidy and evolution in wild and cultivated Dahlia species. Ann Bot 81:647-656. https://doi.org/10.1006/anbo.1998.0614

Geissman TA (1942) Anthochlor pigments. III. The pigments of *Cosmos sulphureus*. J Amer Chem Soc 64:1704–1707. https://doi.org/10.1021/ja01259a066

Geissman TA, Heaton CD (1943) Anthochlor pigments. IV. The pigments of *Coreopsis grandiflora*, Nutt. I. J Amer Chem Soc 65:677–683. https://doi.org/10.1021/ja01244a050

Grabherr MG, Haas BJ, Yassour M et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883

Harborne JB, Greenham J, Eagles J (1990) Malonylated chalcone glycosides in Dahlia.

750 Phytochemistry 29:2899–2900. https://doi.org/10.1016/0031-9422(90)87101-Y
751 Horsch RB, Fry J, Hoffmann N, Neidermeyer J, Rogers SG, Fraley RT (1989) Leaf disc
752 transformation. In: Gelvin SB, Schilperoort RA, Verma DPS, eds. Plant Molecular
753 Biology Manual, pp 63–71. https://doi.org/10.1007/978-94-009-0951-9_5
754 Hoshino A, Mizuno T, Shimizu K, Mori S, Fukada-Tanaka S, Furukawa K, Ishiguro K,
755 Tanaka Y, Iida S (2019) Generation of yellow flowers of the Japanese morning glory
756 by engineering its flavonoid biosynthetic pathway toward aurones. Plant Cell
757 Physiol 60:1871–1879. https://doi.org/10.1093/pcp/pcz101
758 Jez JM, Flynn TG, Penning TM (1997) A new nomenclature for the aldo-keto reductase
759 superfamily. Biochem Pharmacol 54:639–647. https://doi.org/10.1016/S0006-
760 2952(97)84253-0
761 Jez JM, Penning TM (2001) The aldo-keto reductase (AKR) superfamily: an update.
762 Chem Biol Interact 130–132:499–525. https://doi.org/10.1016/S0009-
763 2797(00)00295-7
764 Joung J, Kasthuri GM, Park J, Kang W, Kim H, Yoon B, Joung H, Jeon J (2003) An
765 overexpression of chalcone reductase of *Pueraria montana* var. *lobata* alters
766 biosynthesis of anthocyanin and 5′-deoxyflavonoids in transgenic tobacco. Biochem
767 Biophys Res Commun 303:326–331. https://doi.org/10.1016/S0006-
768 291X(03)00344-9
769 Kaintz C, Molitor C, Thill J, Kampatsikas I, Michael C, Halbwirth H, Rompel A (2014)
770 Cloning and functional expression in *E. coli* of a polyphenol oxidase transcript from
771 *Coreopsis grandiflora* involved in aurone formation. FEBS Letters 588:3417–3426.
772 https://doi.org/10.1016/j.febslet.2014.07.034
773 Lawrence WJC (1931) The genetics and cytology of *Dahlia variabilis*. J Genet 24:257–
774 306

775 Mameda R, Waki T, Kawai Y, Takahashi S, Nakayama T (2018) Involvement of chalcone
776 reductase in the soybean isoflavone metabolon: identification of GmCHR5, which
777 interacts with 2-hydroxyisoflavanone synthase. Plant J 96:56–74.
778 https://doi.org/10.1111/tpj.14014
779 Miosic S, Knop K, Hölscher D, Greiner J, Gosch C, Thill J, Kai M, Shrestha BK,
780 Schneider B, Crecelius AC, Schubert US, Svatos A, Stich K, Halbwirth H (2013) 4-
781 Deoxyaurone formation in *Bidens ferulifolia* (Jacq.) DC. PLoS ONE 8:e61766.
782 https://doi.org/10.1371/journal.pone.0061766
783 Molitor C, Mauracher SG, Pargan S, Mayer RL, Halbwirth H, Rompel A (2015) Latent
784 and active aurone synthase from petals of *C. grandiflora*: a polyphenol oxidase with
785 unique characteristics. Planta 242:519–537. https://doi.org/10.1007/s00425-015-
786 2261-0
787 Molitor C, Mauracher SG, Rompel A (2016) Aurone synthase is a catechol oxidase with
788 hydroxylase activity and provides insights into the mechanism of plant polyphenol
789 oxidases. Proc Natl Acad Sci USA 113:E1806–E1815.
790 https://doi.org/10.1073/pnas.1523575113
791 Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamurai M, Niwa Y, Toyoda K, Matsuoka
792 K, Jimbo T, Kumura T (2007) Development of series of gateway binary vectors,
793 pGWBs, for realizing efficient construction of fusion genes for plant transformation.
794 J Biosci Bioeng 103:34–41. https://doi.org/10.1263/jbb.104.34
795 Ohno S, Hosokawa M, Hoshino A, Kitamura Y, Morita Y, Park K, Nakashima A, Deguchi
796 A, Tatsuzawa F, Doi M, Iida S, Yazawa S (2011a) A bHLH transcription factor,
797 *DvIVS*, is involved in regulation of anthocyanin synthesis in dahlia (*Dahlia*
798 *variabilis*). J Exp Bot 62:5105–5116. https://doi.org/10.1093/jxb/err216

799 Ohno S, Hosokawa M, Kojima M, Kitamura Y, Hoshino A, Tatsuzawa F, Doi M, Yazawa
800 S (2011b) Simultaneous post-transcriptional gene silencing of two different
801 chalcone synthase genes resulting in pure white flowers in the octoploid dahlia.
802 Planta 234:945–958. https://doi.org/10.1007/s00425-011-1456-2
803 Ohno S, Deguchi A, Hosokawa M (2013a) Genetic control of anthocyanin synthesis in
804 dahlia (*Dahlia variabilis*). In: Ramawat KG, Merillon JM, eds. Bulbous plants:
805 Biotechnology, CRC Press, Boca Raton, pp 228–247
806 Ohno S, Deguchi A, Hosokawa M, Tatsuzawa F, Doi M (2013b) A basic helix-loop-helix
807 transcription factor *DvIVS* determines flower color intensity in cyanic dahlia
808 cultivars. Planta 238:331–343. https://doi.org/10.1007/s00425-013-1897-x
809 Ohno S, Hori W, Hosokawa M, Tatsuzawa F, Doi M (2018a) Post-transcriptional
810 silencing of *chalcone synthase* is involved in phenotypic lability in petals and leaves
811 of bicolor dahlia (*Dahlia variabilis*) ‘Yuino’. Planta 247:413–428.
812 https://doi.org/10.1007/s00425-017-2796-3
813 Ohno S, Hori W, Hosokawa M, Tatsuzawa F, Doi M (2018b) Identification of flavonoids
814 in leaves of a labile bicolor flowering dahlia (*Dahlia variabilis*) 'Yuino'. Hort J
815 87:140–148. https://doi.org/10.2503/hortj.OKD-099
816 Ohno S, Ueno M, Doi M (2020) Difference in the *CaMYBA* genome among anthocyanin
817 pigmented cultivars and non-pigmented cultivars in pepper (*Capsicum annuum*).
818 Hort J 89:30–36. https://doi.org/10.2503/hortj.UTD-097
819 Ohno S, Yokota M, Yamada H, Tatsuzawa F, Doi M (2021) Identification of chalcones
820 and their contribution to yellow coloration in dahlia (*Dahlia variabilis*) ray florets.
821 Hort J 90:450–459. https://doi.org/10.2503/hortj.UTD-305
822 Ono E, Fukuchi-Mizutani M, Nakamura N, Fukui Y, Yonekura-Sakakibara K, Yamaguchi
823 M, Nakayama T, Tanaka T, Kusumi T, Tanaka Y (2006) Yellow flowers generated by

824 expression of the aurone biosynthetic pathway. Proc Natl Acad Sci USA 103:11075–
825 11080. https://doi.org/10.1073/pnas.0604246103
826 Pattanaik S, Kong Q, Zaitlin D, Werkman JR, Xie CH, Patra B, Yuan L (2010) Isolation
827 and functional characterization of a floral tissue-specific R2R3 MYB regulator from
828 tobacco. Planta 231:1061–1076. https://doi.org/10.1007/s00425-010-1108-y
829 Polturak G, Breitel D, Grossman N, Sarrion-Perdigones A, Weithorn E, Pliner M, Orzaez
830 D, Granell A, Rogachev I, Aharoni A (2016) Elucidation of the first committed step
831 in betalain biosynthesis enables the heterologous engineering of betalain pigments
832 in plants. New Phytol 210:269–283. https://doi.org/10.1111/nph.13796
833 Price JR (1939) The yellow colouring matter of *Dahlia variabilis*. J Chem Soc 1017–
834 1018. https://doi.org/10.1039/JR9390001017
835 Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing
836 phylogenetic trees. Mol Biol Evol 4:406-425.
837 https://doi.org/10.1093/oxfordjournals.molbev.a040454
838 Sato-Masumoto N, Ito M (2014) Two types of alcohol dehydrogenase from *Perilla* can
839 form citral and perillaldehyde. Phytochemistry 104:12–20.
840 https://doi.org/10.1016/j.phytochem.2014.04.019
841 Schlangen K, Miosic S, Halbwirth, H (2010a) Allelic variants from *Dahlia variabilis*
842 encode flavonoid 3′-hydroxylases with functional differences in chalcone 3-
843 hydroxylase activity. Arch Biochem Biophys 494:40–45.
844 https://doi.org/10.1016/j.abb.2009.11.015
845 Schlangen K, Miosic S, Thill J Halbwirth H (2010b) Cloning, functional expression, and
846 characterization of a chalcone 3-hydroxylase from *Cosmos sulphureus*. J Exp Bot
847 61:3451–3459. https://doi.org/10.1093/jxb/erq169
848 Semwal RB, Semwal DK, Combrinck S, Viljoen A (2015) Butein: From ancient

traditional remedy to modern nutraceutical. Phytochem Lett 11:188–201. https://doi.org/10.1016/j.phytol.2014.12.014

Shimada N, Nakatsuka T, Nishihara M, Yamamura S, Ayabe S, Aoki T (2006) Isolation and characterization of a cDNA encoding polyketide reductase in *Lotus japonicus*. Plant Biotechnol 23:509–513. https://doi.org/10.5511/plantbiotechnology.23.509

Sun L, Ruppert M, Sheludko Y, Warzecha H, Zhao Y, Stöckigt J (2008) Purification, cloning, functional expression and characterization of perakine reductase: The first example from the AKR enzyme family, extending the alkaloidal network of the plant Rauvolfia. Plant Mol Biol 67:455–467. https://doi.org/10.1007/s11103-008-9331-7

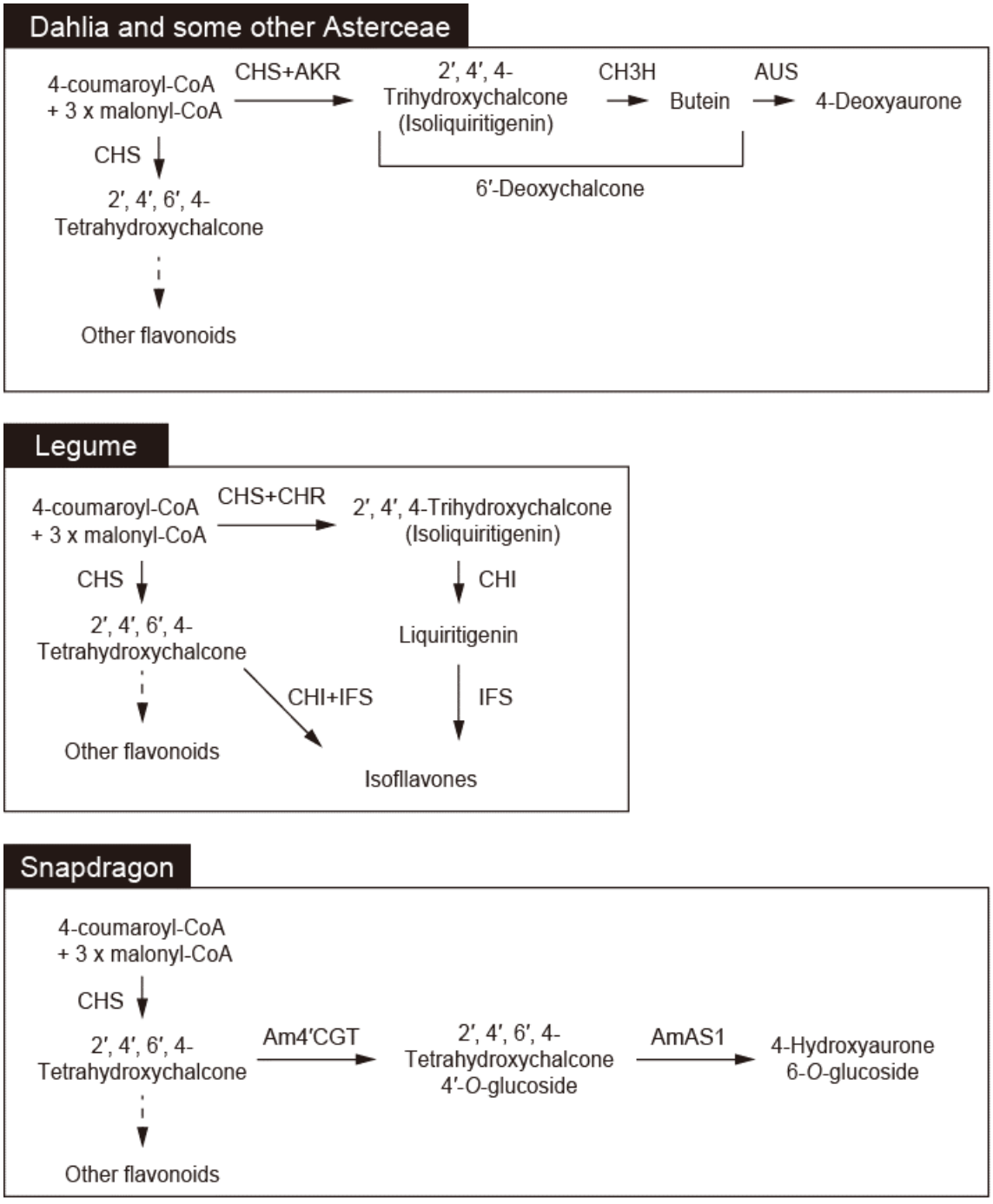
Sun L, Chen Y, Rajendran C, Mueller U, Panjikar A, Wang M, Mindnich R, Rosenthal C, Penning TM, Stöckigt J (2012) Crystal structure of Perakine Reductase, founding member of a novel aldo-keto reductase (AKR) subfamily that undergoes unique conformational changes during NADPH binding. J Biol Chem 287:11213–11221. https://doi.org/10.1074/jbc.M111.335521

Suzuki H, Nakayama T, Yonekura-Sakakibara K, Fukui Y, Nakamura N, Yamaguchi MA, Tanaka Y, Kusumi T, Nishino T (2002) cDNA cloning, heterologous expressions, and functional characterization of malonyl-coenzyme A: anthocyanidin 3-*O*-glucoside-6′-*O*-malonyltransferase from dahlia flowers. Plant Physiol 130:2142–2151. https://doi.org/10.1104/pp.010447

Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 38:3022–3027. https://doi.org/10.1093/molbev/msab120

Walliser B, Lucaciu CR, Molitor C, Marinovic S, Nitarska DA, Aktaş D, Rattei T, Kampatsikas I, Stich K, Haselmair-Gosch C, Halbwirth H (2021) Dahlia variabilis cultivar ‘Seattle’ as a model plant for anthochlor biosynthesis. Plant Physiol

874 Biochem 159:193–201. https://doi.org/10.1016/j.plaphy.2020.12.016
875 Welle R, Grisebach H (1988) Isolation of a novel NADPH-dependent reductase which
876 coacts with chalcone synthase in the biosynthesis of 6′-deoxychalcone. FEBS Let
877 263:191–198. https://doi.org/10.1016/0014-5793(88)80318-1
878 Welle R, Schroder G, Schiltz E, Grisebach H, Schroder J (1991) Induced plant responses
879 to pathogen attack: analysis and heterologous expression of the key enzyme in the
880 biosynthesis of phytoalexins in soybean (*Glycine max* L. Merr. cv. Harosoy 63). Eur
881 J Biochem 196:423–430 https://doi.org/10.1111/j.1432-1033.1991.tb15833.x
882 Zhang Z, Li DW, Jin JH, Yin YX, Zhang HX, Chai WG, Gong ZH (2015) VIGS approach
883 reveals the modulation of anthocyanin biosynthetic genes by *CaMYB* in chili pepper
884 leaves. Front Plant Sci 6:500. https://doi.org/10.3389/fpls.2015.00500



885

886

887 **Fig. 1** Simplified illustration of chalcone and aurone biosynthesis pathways in different
888 plants. Abbreviations: AKR, Aldo-keto reductase; AmAS1, aureusidin synthase;
889 Am4′CGT, chalcone 4′-*O*-glucosyltransferase; AUS, aurone synthase; CH3H, chalcone
890 3-hydroxylase; CHI, chalcone isomerase; CHR, chalcone reductase; CHS, chalcone
891 synthase; IFS, isoflavone synthase.

892

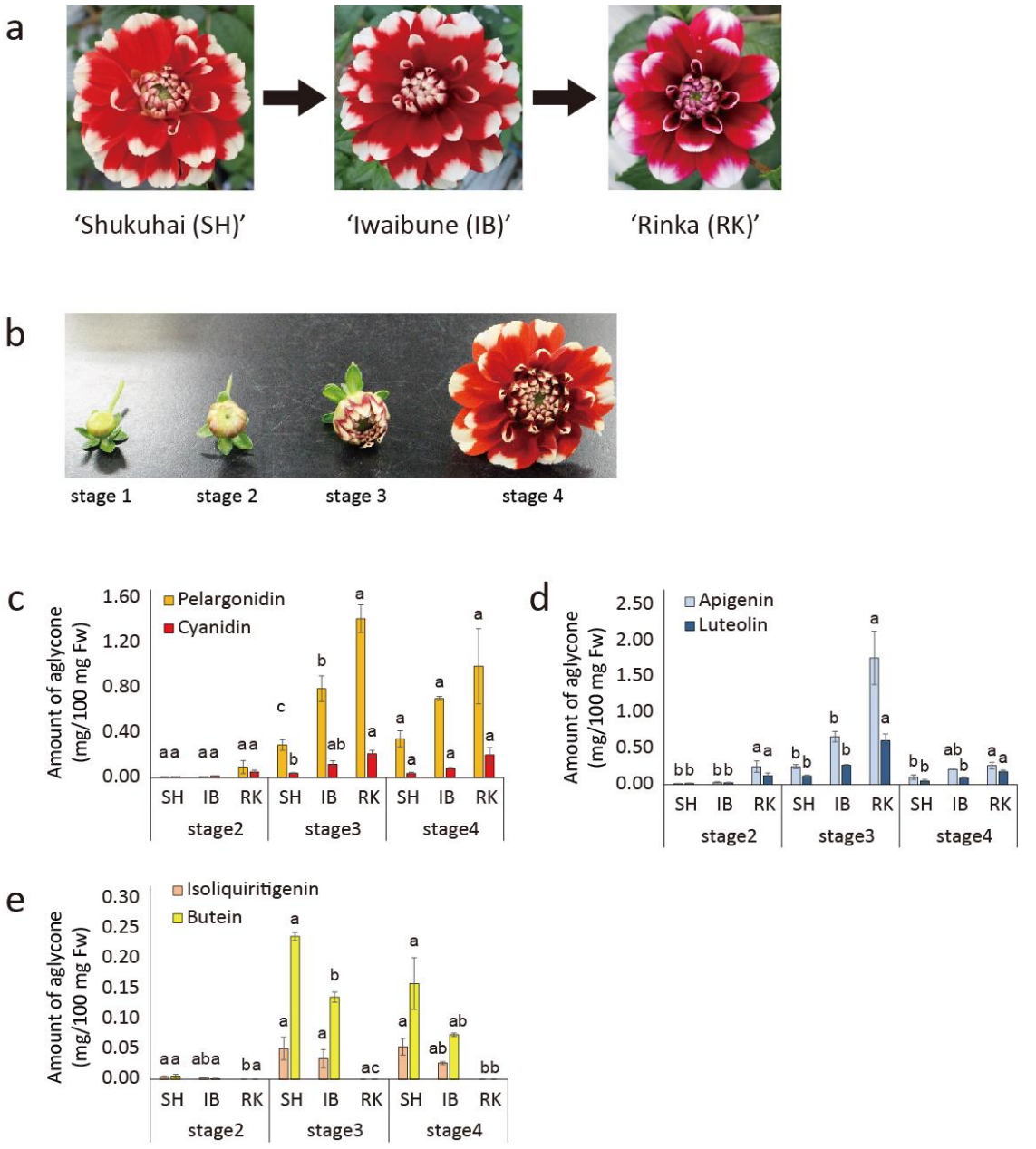


Fig. 2 Pigment analysis in 'SH,' 'IB,' and 'RK'. **a** Phenotypes of dahlia cultivar used in this study. **b** Ray floret at different developmental stages used in this study. **c-e** Pigment content in ray florets at developmental stages 2 to 4. **c** Anthocyanidin content. **d** flavone content. **e** 6'-deoxychalcone content. Different letters above the bars indicate significant differences $P < 0.05$ ($n = 3$).

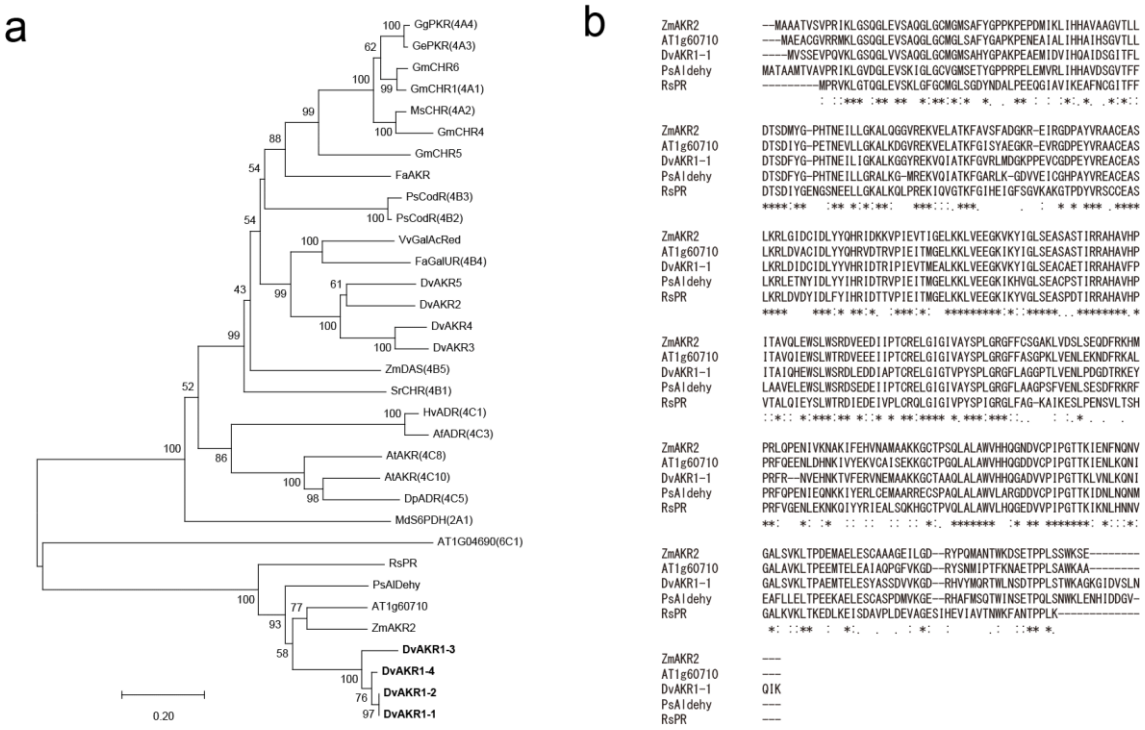


Fig. 3 Phylogenetic analysis of DvAKR1. **a** Phylogenetic tree for plant Aldo-keto reductase superfamily protein. Full amino acid sequences were aligned using ClustalW, and the tree was constructed by the neighbor-joining method. Bootstrap values obtained with 1000 repetitions are indicated on each branch. **b** Comparison of putative amino acid sequence among *DvAKR1-1* with its homologs.

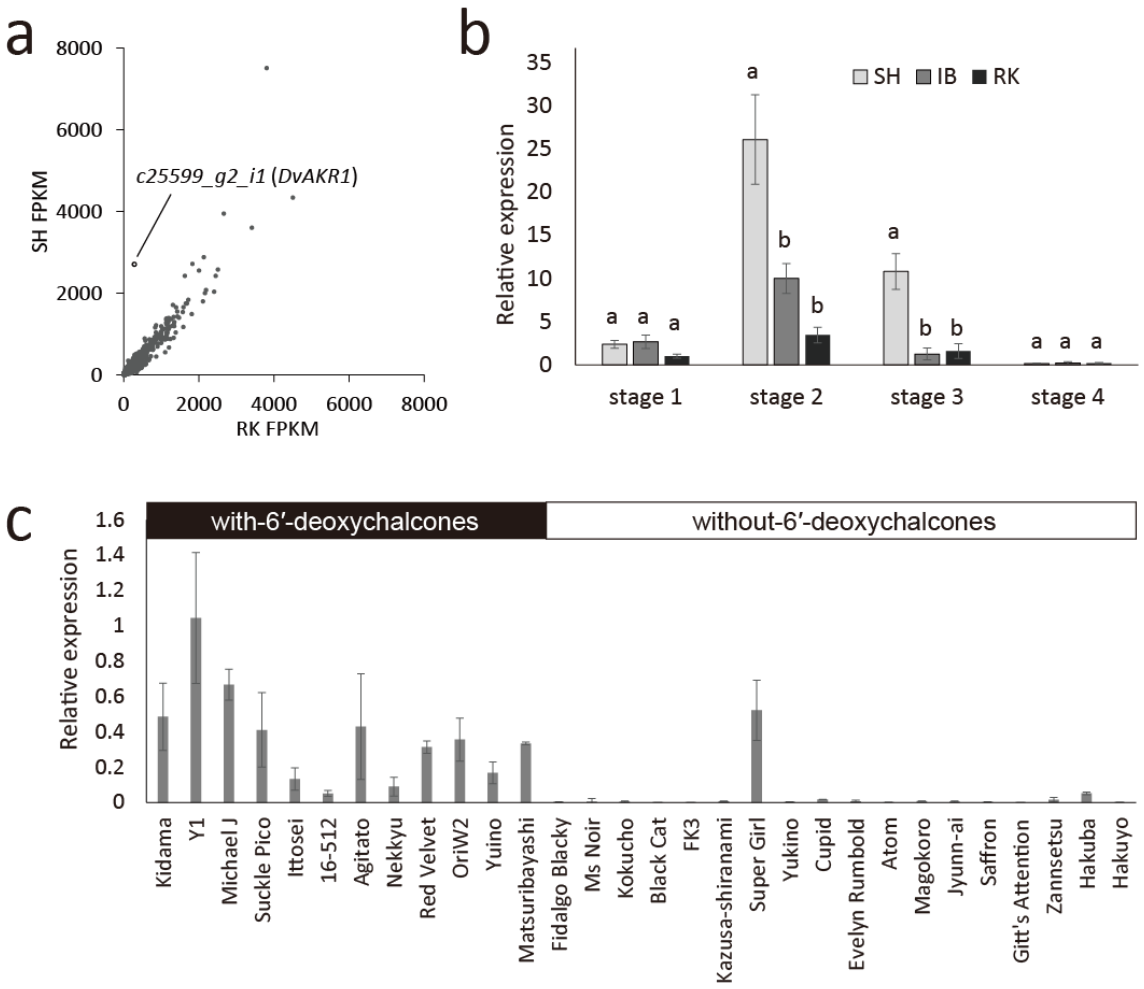


Fig. 4 *DvAKR1* expression analysis. **a** Scatter plot of fragments per kilobase of exon per million mapped reads (FPKM) values between 'SH' and 'RK' from RNA-seq. **b** Relative expression of *DvAKR1* in ray florets of 'SH,' 'IB,' and 'RK' at different developmental stages. Different letters above the bars indicate significant differences $P < 0.05$ ($n = 3$). **c** Relative expression of *DvAKR1* in ray florets of 30 cultivars or seedling lines tested in this study. *DvActin* was used as an internal control. Bars represent standard error ($n = 3$).

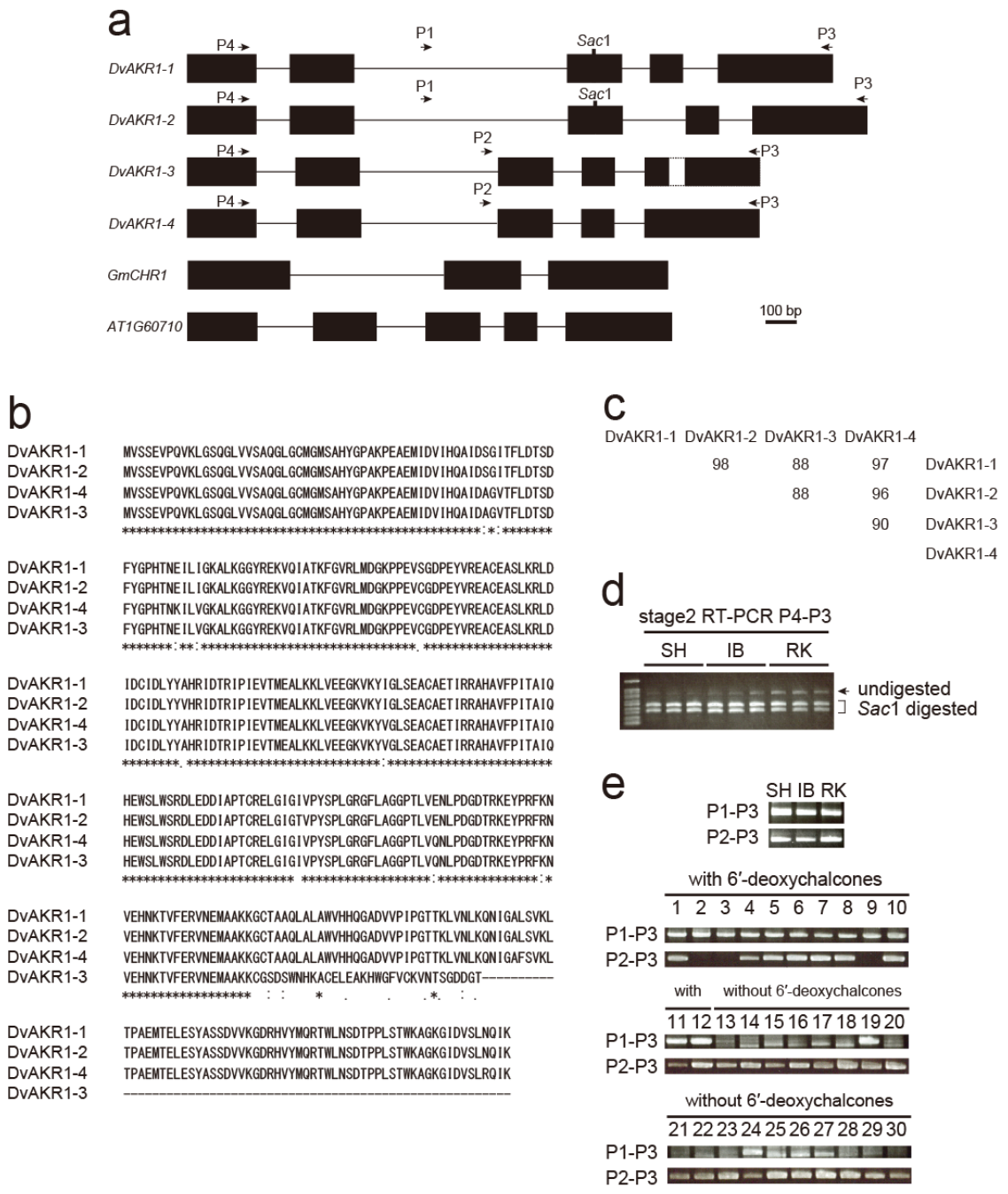


Fig. 5 Genetic analysis of the candidate *DvAKR1* gene. **a** Genetic structure of *DvAKR1*. Black boxes and horizontal lines indicate exons and introns, respectively. **b** Alignment of putative DvAKR1-1, DvAKR1-2, DvAKR1-3, and DvAKR1-4 amino acid sequences. **c** Percentage amino acid identity among DvAKR1-1, DvAKR1-2, DvAKR1-3, and DvAKR1-4. **d** RT-PCR analysis of stage 2 ray florets using P4 and P3 primers, and *Sac*I restriction enzyme digestion. Sequences of *DvAKR1-1* and *DvAKR1-2* but not *DvAKR1-3* and *DvAKR1-4* harbored *Sac*I restriction site. **e** Genomic PCR analysis of *DvAKR1* in the cultivars accumulating and lacking 6'-deoxychalcones. 1: 'Kidama' (yellow), 2: 'Y1'

(yellow), 3: ‘Michael J’ (red variegation on yellow), 4: ‘Suckle Pico’ (yellow with red flush), 5: ‘Ittosei’ (pale yellow), 6: ‘16-512’ (pale yellow), 7: ‘Agitato’ (red), 8: ‘Nekkyu’ (red), 9: ‘Red Velvet’ (red), 10: ‘OriW2’ (red-white), 11: ‘Yuino’ (red-white), 12: ‘Matsuribayashi’ (red-white), 13: ‘Fidalgo Blacky’ (black), 14: ‘Ms. Noir’ (black), 15: ‘Kokucho’ (black), 16: ‘Black Cat’ (black), 17: ‘FK3’ (black), 18: ‘Kazusa-shiranami’ (black-white), 19: ‘Super Girl’ (deep purple), 20: ‘Yukino’ (deep purple), 21: ‘Cupid’ (purple), 22: ‘Evelyn Rumbold’ (purple), 23: ‘Atom’ (purple), 24: ‘Magokoro’ (pink), 25: ‘Jun-ai’ (pink), 26: ‘Saffron’ (pink), 27: ‘Gitt’s Attention’ (ivory white), 28: ‘Zannsetsu’ (ivory white), 29: ‘Hakuba’ (ivory white), 30: ‘Hakuyo’ (ivory white).

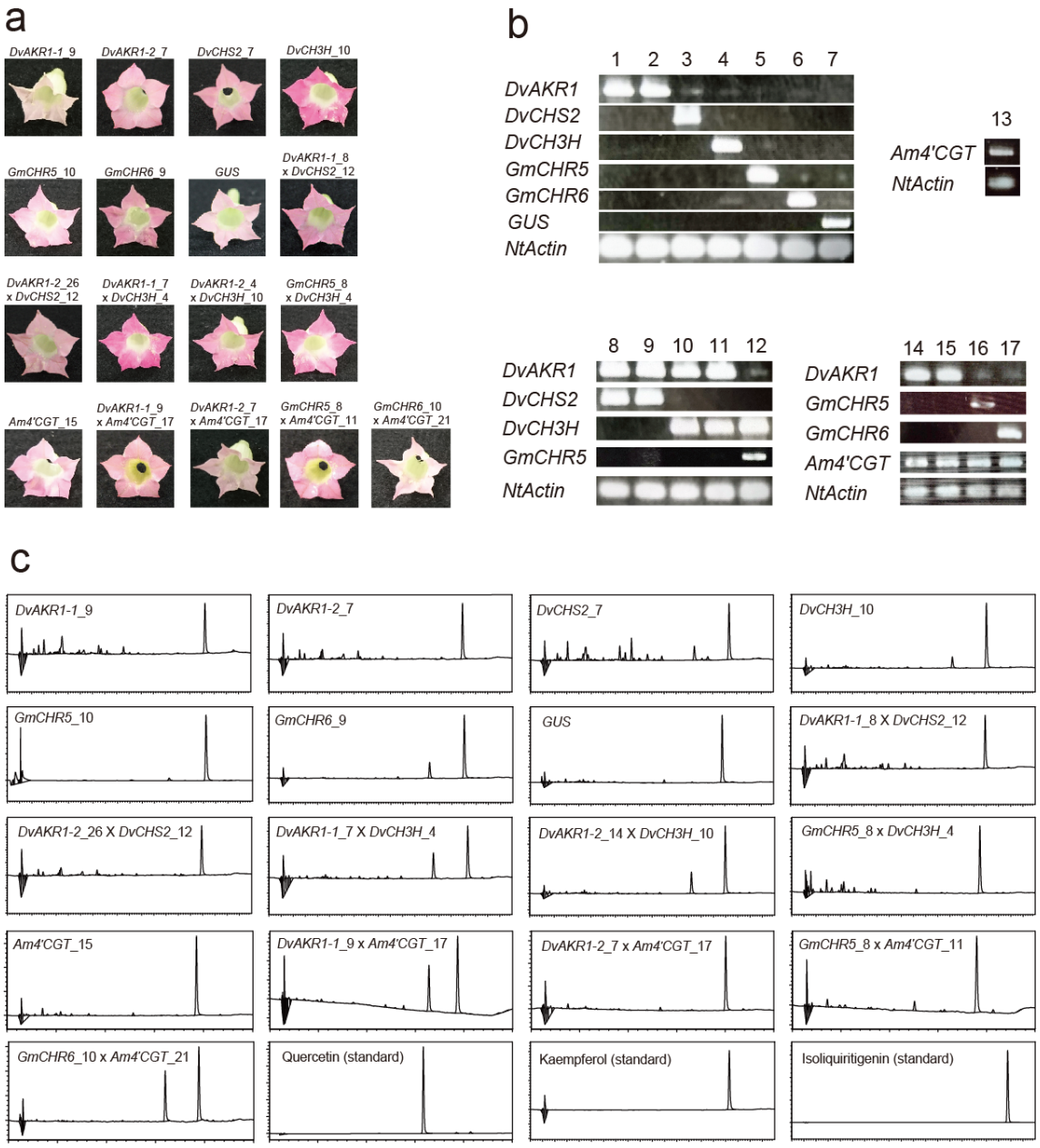


Fig. 6 Over-expression analysis in transgenic *N. tabacum* plants. **a** Phenotypes of transgenic over-expression flowers. **b** RT-PCR validation of transgenic expression in flowers. 1: *DvAKR1-1*_9, 2: *DvAKR1-2*_7, 3: *DvCHS2*_7, 4: *DvCH3H*_10, 5: *GmCHR5*_10, 6: *GmCHR6*_9, 7: *GUS*, 8: *DvAKR1-1*_8 x *DvCHS2*_12, 9: *DvAKR1-2*_26 x *DvCHS2*_12, 10: *DvAKR1-1*_7 x *DvCH3H*_4, 11: *DvAKR1-2*_14 x *DvCH3H*_10, 12: *GmCHR5*_8 x *DvCH3H*_4, 13: *Am4'CGT*_15, 14: *DvAKR1-1*_9 x *Am4'CGT*_17, 15: *DvAKR1-2*_7 x *Am4'CGT*_17, 16: *GmCHR5*_8 x *Am4'CGT*_11, 17: *GmCHR6*_10 x *Am4'CGT*_21. **c** HPLC chromatograms of transgenic over-expression flowers analyzed at 380 nm.

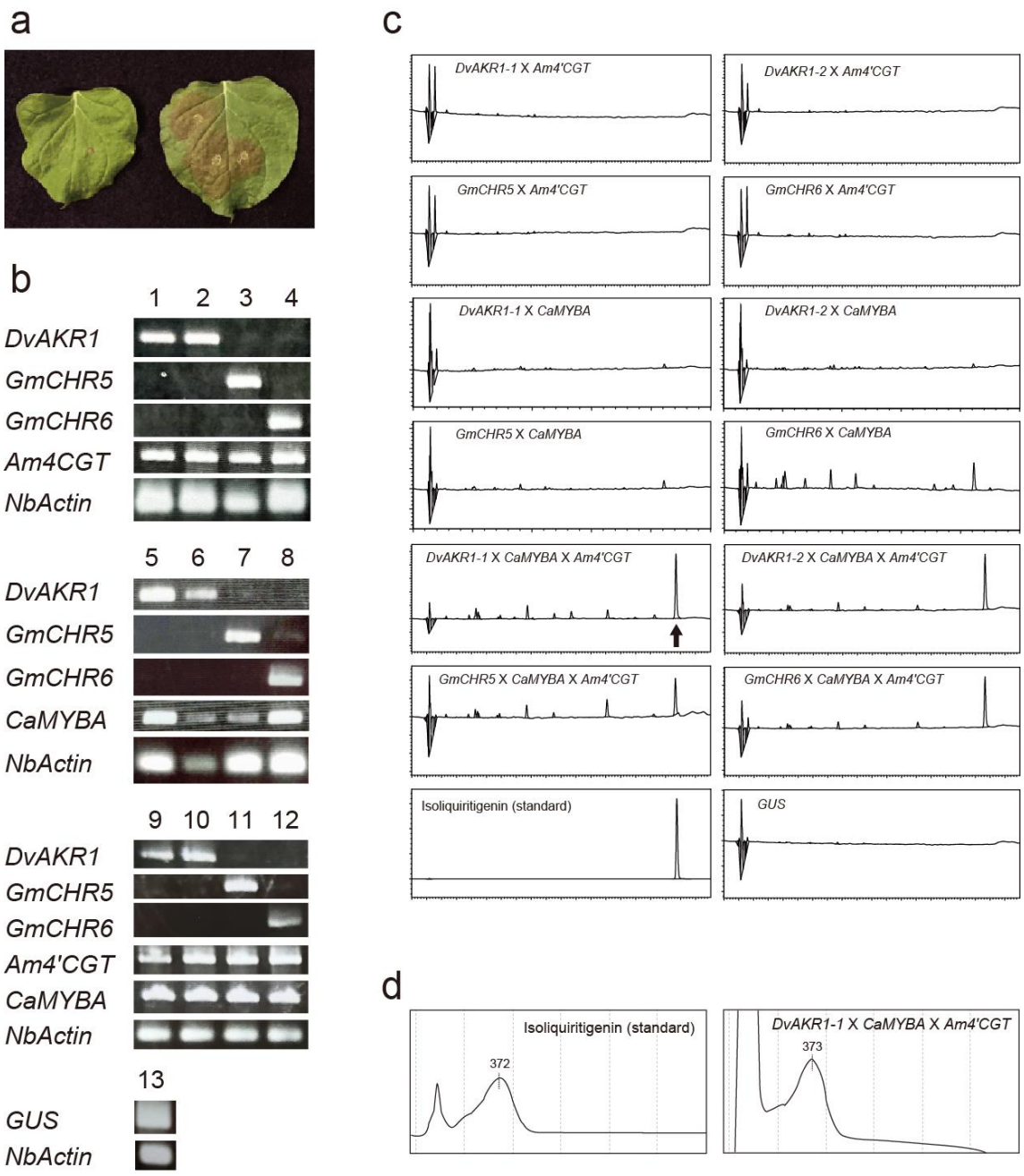


Fig. 7 Transient over-expression analysis by agroinfiltration in leaves. **a** Phenotypes of infiltrated leaves. **b** RT-PCR validation of transgenic over-expression in leaves. 1: *DvAKR1-1* x *Am4'CGT*, 2: *DvAKR1-2* x *Am4'CGT*, 3: *GmCHR5* x *Am4'CGT*, 4: *GmCHR6* x *Am4'CGT*, 5: *DvAKR1-1* x *CaMYBA*, 6: *DvAKR1-2* x *CaMYBA*, 7: *GmCHR5* x *CaMYBA*, 8: *GmCHR6* x *CaMYBA*, 9: *DvAKR1-1* x *CaMYBA* x *Am4'CGT*, 10: *DvAKR1-2* x *CaMYBA* x *Am4'CGT*, 11: *GmCHR5* x *CaMYBA* x *Am4'CGT*, 12: *GmCHR6* x *CaMYBA* x *Am4'CGT*. **c** HPLC chromatograms of over-expression leaves analyzed at 380 nm. **d** Photodiode array analysis showing peaks indicated by an arrow in **c** of *DvAKR1-1* x *CaMYBA* x *Am4'CGT* in triple over-expression leaf and isoliquiritigenin standard.

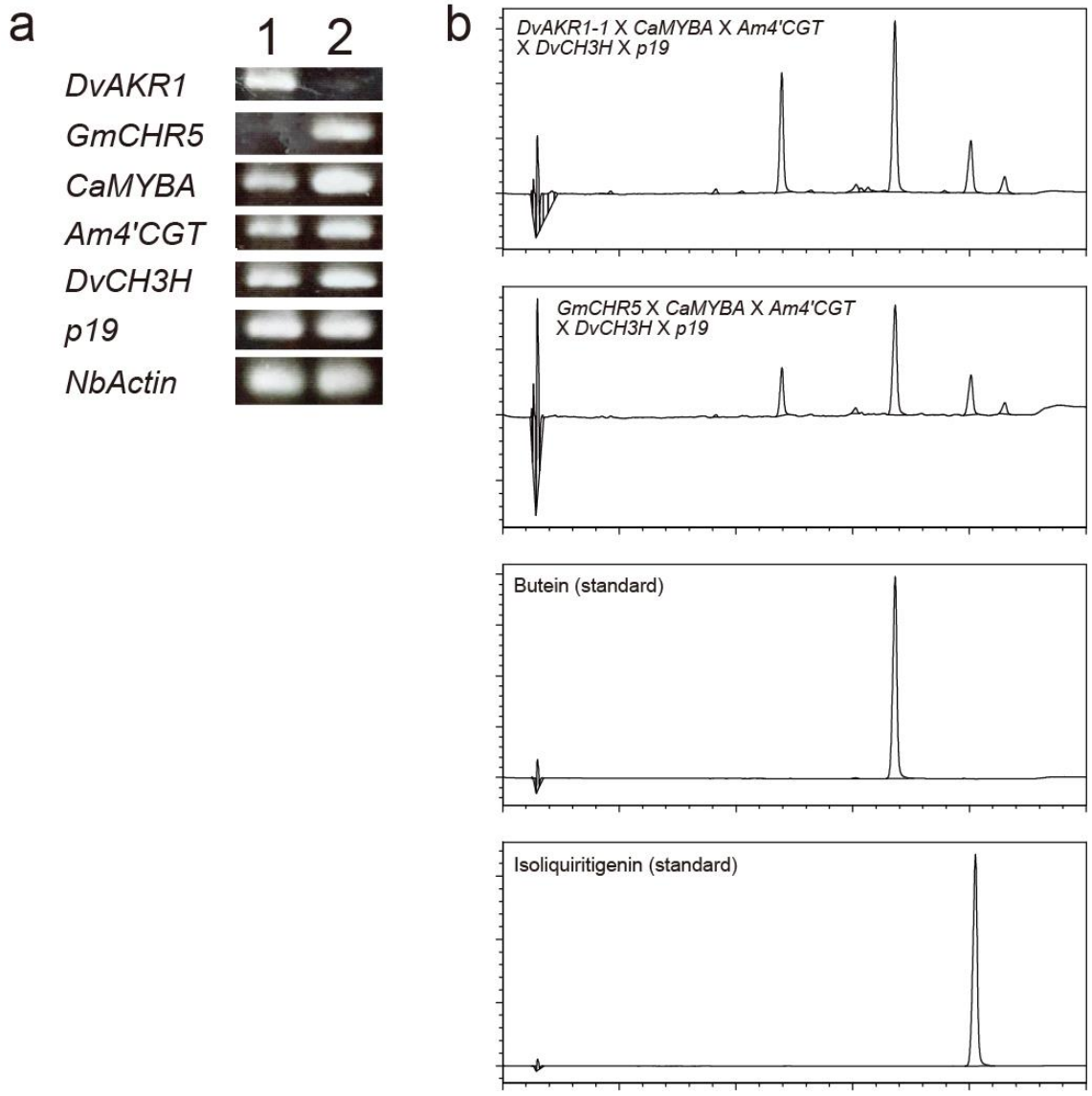


Fig. 8 Production of butein in benthamiana tobacco by agroinfiltration. **a** RT-PCR validation of transgenic expression in leaves. 1: *DvAKR1-1* x *CaMYBA* x *Am4'CGT* x *DvCH3H* x *p19*, 2: *GmCHR5* x *CaMYBA* x *Am4'CGT* x *DvCH3H* x *p19*. **b** HPLC chromatograms analyzed at 380 nm.

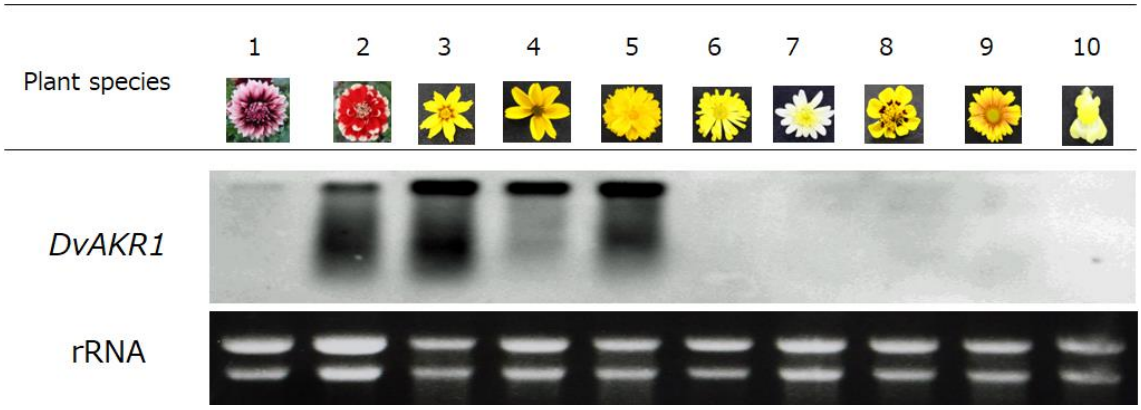


Fig. 9 Hetero-probe RNA gel blot analysis using *DvAKR1* CDS probe. 1: *D. variabilis* 'Kazusa-shiranami', 2: *D. variabilis* 'Shukuhai', 3: *Coreopsis grandiflora* 'Fairy Golden', 4: *Bidens ferulifolia* 'JuJu Gold', 5: *Cosmos sulphureus*, 6: *Chrysanthemum morifolium* 'Laub Fusha', 7: *Argyranthemum frutescens* 'Lemon Yellow', 8: *Tagetes patula* 'Harlequin', 9: *Gaillardia* × *grandiflora* 'Arizona Apricot' 10: *Antirrhinum majus*.