



Diversity of genetic lesions characterizes new *Arabidopsis* flavonoid pigment mutant alleles from T-DNA collections

Nan Jiang, Yun Sun Lee, Eric Mukundi, Fabio Gomez-Cano, Luz Rivero¹, Erich Grotewold^{*,1}

Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824-6473, USA

ARTICLE INFO

Keywords:

Transparent Testa
Flavonoid
Proanthocyanidin
Anthocyanin
Forward genetics
T-DNA insertion

ABSTRACT

The visual phenotypes afforded by flavonoid pigments have provided invaluable tools for modern genetics. Many *Arabidopsis transparent testa* (*tt*) mutants lacking the characteristic proanthocyanidin (PA) seed coat pigmentation and often failing to accumulate anthocyanins in vegetative tissues have been characterized. These mutants have significantly contributed to our understanding of flavonoid biosynthesis, regulation, and transport. A comprehensive screening for *tt* mutants in available large T-DNA collection lines resulted in the identification of 16 independent lines lacking PAs and anthocyanins, or with seed coat pigmentation clearly distinct from wild type. Segregation analyses and the characterization of second alleles in the genes disrupted by the indexed T-DNA insertions demonstrated that all the lines contained at least one additional mutation responsible for the *tt* phenotypes. Using a combination of RNA-Seq and whole genome re-sequencing and confirmed through complementation, we show here that these mutations correspond to novel alleles of *ttg1* (two alleles), *tt3* (two alleles), *tt5* (two alleles), *ban* (two alleles), *tt1* (two alleles), and *tt8* (six alleles), which harbored additional T-DNA insertions, indels, missense mutations, and large genomic deletion. Several of the identified alleles offer interesting perspectives on flavonoid biosynthesis and regulation.

1. Introduction

Flavonoid formation continues to be one of the best characterized plant specialized metabolite biosynthesis pathways [1–3]. Over the past three decades, *Arabidopsis transparent testa* (*tt*) mutants lacking proanthocyanidins (PAs) in the seed coat, and sometimes deficient in anthocyanins in vegetative tissues, have significantly contributed to advancing our understanding of the biosynthesis, regulation and storage of flavonoid compounds [4,5]. Today, there are 19 known loci in *Arabidopsis* that encode products required for seed coat PA accumulation, and several also affect anthocyanin accumulation [3,6].

The first committed step in the flavonoid biosynthesis pathway is catalyzed by chalcone synthase (CHS) [5,7], encoded by *TT4* (At5g13930, Supplementary Fig. S1 and Supplementary Table S1). The resulting naringenin chalcone is the substrate for a chalcone isomerase (CHI) [5,8], encoded by *TT5* (At1g51670), which converts chalcone into a flavanone (e.g., naringenin). Flavanones can be first hydroxylated in the B-ring by a flavanone 3-hydroxylase (F3H, encoded by *TT6*, At3g51240) [9,10] and then in the C-ring by a flavonoid 3'-hydroxylase (F3'H, encoded by *TT7*, At5g07990) [11]. The resulting 3-hydroxy

flavanones can be converted into flavonols by the action of flavonol synthases (FLS1 and FLS3, At5g08640 and At5g63590) [12,13], or reduced to flavan-3,4-diols (leucoanthocyanidins) by the action of dihydroflavonol 4-reductase (DFR, encoded by *TT3*, At5g42800) [8,14]. Leucoanthocyanidins are then oxidized by the action of a leucoanthocyanidin dioxygenase (LDOX, a.k.a anthocyanidin synthase, ANS, At4g22880) to 3-hydroxyanthocyanidins [15–17]. In *Arabidopsis* seed coats, 3-hydroxyanthocyanidins serve as the precursors for the formation of flavan-3-ol subunits by the catalytic reaction of anthocyanidin reductase (ANR, encoded by *BANYLUS*, BAN, At1g61720) [18–20]. The flavan-3-ol subunits (such as catechins and epicatechins) are then believed to be transported to the vacuole and polymerized to form PAs (or condensed tannins) that provide the characteristic brown color to seeds with the involvement of a membrane trafficking factor GFS9 (*TT9*, At3g28430) [21], the glutathione *S*-transferase *phi* 12 (*TT19*, At5g17220) [22], the multi-antimicrobial extrusion protein (MATE) transporter (*TT12*, At3g59030) [23,24], the tonoplast P_{3A}-ATPase AHA10 (*TT13*, At1g17260) [25,26], and the laccase-like enzyme LAC15 (*TT10*, At5g48100) [27]. However, in *Arabidopsis* vegetative tissues, particularly under stress conditions, anthocyanidins undergo a series of

* Corresponding author.

E-mail address: grotewol@msu.edu (E. Grotewold).

¹ Former Affiliation: Arabidopsis Biological Resource Center (ABRC), Department of Molecular Genetics and Center for Applied Plant Sciences (CAPS), The Ohio State University, Columbus, OH 43210, USA.

<https://doi.org/10.1016/j.plantsci.2019.110335>

Received 28 May 2019; Received in revised form 17 November 2019; Accepted 18 November 2019

Available online 26 November 2019

0168-9452/ © 2019 Elsevier B.V. All rights reserved.

glycosylations and other modifications, resulting in the formation of more than 14 different kinds of anthocyanins [28–31].

The regulation of *Arabidopsis* flavonoids is also well described [32]. As in other plants, the interplay between R2R3-MYB and basic helix-loop-helix (bHLH) transcription factors with the participation of a WD protein is essential for both anthocyanin and PA accumulation, although the specific members of each of these groups can vary [33–36]. For example, PA accumulation is largely controlled by TT2 (MYB123, At5g35550) [37] and TT8 (bHLH042, At4g09820) [38], with the participation of TTG1 (TRANSPARENT TESTA GLABRA1, At5g24520), a WD protein that also participates in the control of trichomes and root hairs [39–41]. Meanwhile, WRKY (TTG2/WRKY44, At2g37260) [42,43] and WIP (TT1/WIP1, At1g34790) [44,45] transcription factors, an UDP-glucose:sterol glucosyltransferase (TT15/UGT80B1, At1g43620) [46,47], and a MADS-box transcription factor (TT16/AGL32, At5g23260) [47–49] also contribute to PA accumulation and deposition. Anthocyanin regulation is controlled by the combinatorial action of PAP1/MYB75 (At1g56650) and related R2R3-MYB genes PAP2/MYB90 (At1g66390), MYB113 (At1g66370), and MYB114 (At1g66380) [50,51], working in concert with GL3 (bHLH001, At5g41315) and EGL3 (bHLH002, At1g63650), two bHLH transcription factors that also participate in trichome formation [52,53]. Previous studies also indicated the participation of TT8 in anthocyanin accumulation [53], involving a MAPK cascade activated by sucrose signaling [54]. TTG1 participates in all these bHLH-MYB interactions [55]. Other anthocyanin regulators have been described, including the small MYB proteins MYBL2 (R3-MYB, At1g71030) and MYBD (At1g70000) that negatively control pigmentation [56,57], and the homeodomain protein ANTHOCYANINLESS2 (ANL2, At4g00730) that represses anthocyanin accumulation in subepidermal, but not in epidermal cells [58].

In this study, we asked the question of whether new genes involved in seed coat pigment formation could be identified from the very large indexed T-DNA collections [59] propagated and distributed by the Arabidopsis Biological Resource Center (ABRC). By systematically screening the T-DNA collections propagated by ABRC, we identified 16 independent lines with *transparent testa* (*tt*) phenotypes and annotated T-DNA insertion sites. However, using a combination of second allele characterization and segregation analyses, we determined that the genes harboring the indexed T-DNA insertions in these mutants were not responsible for the *tt* phenotypes. To characterize the causative genes for the *tt* phenotypes, we first investigated whether the 16 mutants also affected anthocyanin accumulation. Through a combination of RNA-Seq analyses on *Arabidopsis* seedlings under anthocyanin inductive conditions (AIC) and whole genome re-sequencing of pooled mutant tissues, we determined that the 16 mutations corresponded to new alleles of *tt1* (two mutant lines), *tt3* (two mutant lines), *tt5* (two mutant lines), *ban* (two mutant lines), *tt8* (six mutant lines), and *ttg1* (two mutant lines). The causative mutations in the known flavonoid genes are diverse, including T-DNA insertions, indels, missense mutations in coding regions, and large deletions of genomic region.

2. Materials and methods

2.1. Chemicals, plant materials, and growth conditions

All the chemicals and HPLC grade solvents were purchased from Sigma. *Arabidopsis thaliana* (L.) accession Columbia-0 (Col-0), 16 mutants with *tt* phenotypes, and second alleles were obtained from the ABRC (Columbus, OH). The information of the newly identified 16 mutants and second alleles used in this study are listed in Table 1 and Supplementary Table S2. All the plants were grown in Suremix growth media (Michigan Grower Products Inc, MI, USA) under controlled conditions (22 °C and 16 h light/8 h dark photoperiod).

2.2. Constructs for complementation

The pENTRY clones for *TTG1* (TOPO-U06-H05) and *TT8* (TOPO-U08-D02) were obtained from the ABRC, and subsequently integrated into the pGWB502 binary vector [60] to make the p35S::TTG1 and p35S::TT8 constructs via LR reactions, which were catalyzed by Gateway® LR Clonase® II enzyme mix (ThermoFisher Scientific, CA, USA) by *in vitro* recombination. The resulting constructs were then introduced into *Agrobacterium tumefaciens* strain (GV3101) and used for transformation as follows: p35S::TTG1 construct into two *ttg1* alleles (*ttg1-23*, SALK_012389 and *ttg1-24*, SAIL_1175_C01) and p35S::TT8 construct into four *tt8* alleles (*tt8-9*, SALK_010879; *tt8-10*, SALK_137974; *tt8-12*, SALK_206210; and *tt8-13*, SAIL_239_C11) by floral dip method [61].

For selection of transformed lines, seeds were surface-sterilized and cultured on Murashige and Skoog (MS) solid media [62] including 5 % sucrose with hygromycin (50 mg/L). For selection, seeds were incubated in 4 °C for two days for stratification and placed at 22 °C with 16 h light/8 h dark photoperiod for germination and growth.

2.3. DMACA staining assay

For the DMACA (*p*-dimethylaminocinnamaldehyde) staining assay, *Arabidopsis* seeds were stained with the reagent (2 % [w/v] DMACA in 3 M HCl with 50 % methanol) for two days at room temperature to characterize spatial patterns of proanthocyanidin accumulation in seed coats [63]. The stained seeds were then washed three times with 70 % ethanol before taking pictures with a Nikon SMZ1500 camera with Nikon WD54 HR Plan Apo lens (Tokyo, Japan).

2.4. Anthocyanin inductive condition

For AIC, we carried out the experiments as previously reported [64]. Briefly, surface-sterilized seeds were plated into water containing 3 % sucrose and incubated two days at 4 °C for stratification. After cold treatment of two days, the seeds were cultured under continuous cool-white light with 20 rpm shaking at 23 °C for four days. For naringenin (Nar) complementation, *Arabidopsis* seedlings were first grown under AIC for three days and then Nar was added to a final concentration of 100 µM for 24 h.

2.5. Anthocyanin extraction and analysis

Arabidopsis seedlings were first grown under AIC with and without Nar for four days (24 h with Nar treatment). Seedlings were then lyophilized and dry tissue weighted. Total anthocyanins were extracted in 50 % methanol containing 3 % formic acid at 50 µg of dry weight per µL of extraction buffer overnight at room temperature. The extracts (20 µL) were analyzed with a 2695 Alliance HPLC system (Waters Corp., Beverly, MA) on a Symmetry C18 column (Waters, 3.5 µm 4.6 × 75 mm) at 35 °C with flow rate 1 mL/min. Buffer A corresponds to 5 % formic acid in water and buffer B to 5 % formic acid in acetonitrile. The gradient running condition was: 0 min, 100 % A and 0 % B; 20 min, 75 % A and 25 % B; 22 min, 20 % A and 80 % B; 22.1 min, 0 % A and 100 % B; 25 min, 0 % A and 100 % B; 25.1 min, 100 % A and 0 % B; and 30 min, 100 % A and 0 % B. Anthocyanin profiles were recorded at 532 nm and individual anthocyanin peaks from the HPLC were collected and applied to the LC-MS/MS to determine anthocyanin composition. The LC-MS/MS running condition and data analysis is adapted from [64].

2.6. Arabidopsis genomic DNA extraction and PCR amplification

Briefly, *Arabidopsis* leaves (~200 mg) were grounded into powder and mixed with 1 mL of urea buffer (7 M of urea, 350 mM of NaCl, 50 mM Tris-HCl pH 8.0, 20 mM EDTA pH 8.0, 1 % w/v N-

Table 1
Summary of RNA-Seq and whole genome re-sequencing identified mutation types for the lines with *tt* phenotypes.

Allele	Gene locus	Germplasm	Mutation type	Mutant phenotypes	Tests to validate allele
<i>ttg1-23</i> ^a	At5g24520	SALK_012389	T-DNA insertion or indel: T-DNA insertion or deletion in exon 1	Yellow seeds, trichomeless	Crosses with <i>ttg1-21</i> (GK-580A05) Complementation with p35S::TTG1 Crosses with <i>ttg1-21</i> (GK-580A05)
<i>ttg1-24</i> ^a	At5g24520	SAIL_1175_C01	Genome deletion: <i>TTG1</i> and four nearby genes (At5g24490, At5g24500, At5g24510, and At5g24530)	Yellow seeds, trichomeless	Complementation with p35S::TTG1 Crosses with <i>tt3-2</i> (GK-295C10)
<i>tt3-4</i> ^a	At5g42800	SAIL_649_G09	Missense mutation: GGA to AGA; Gly to Arg at amino acid 130	Yellow seeds	Crosses with <i>tt3-2</i> (GK-295C10)
<i>tt3-5</i> ^a	At5g42800	SAIL_1180_D08	Indel: 3 bp deletion in exon 5 of <i>TT3</i> at amino acid 233	Yellow seeds	Crosses with <i>tt5-2</i> (GK-176H03)
<i>tt5-4</i> ^a	At3g55120	SAIL_641_E10	T-DNA insertion or indel: T-DNA insertion or deletion in intron 1	Yellow seeds	Crosses with <i>tt5-2</i> (GK-176H03)
<i>tt5-5</i> ^{a,b}	At3g55120	SALK_023403	T-DNA insertion: between 68 and 69 bp of intron 3	Yellow seeds	Crosses with <i>tt5-2</i> (GK-176H03)
<i>ban-5</i> ^{a,b}	At1g61720	SALK_052000	Indel: 27 bp deletion in exon 4	Dark grey seeds	Crosses with <i>ban-4</i> (SALK_040250)
<i>ban-6</i> ^{a,b}	At1g61720	SAIL_874_C01	Indel: 10 bp deletion in exon 4	Dark grey seeds	Crosses with <i>ban-4</i> (SALK_040250)
<i>tt1-5</i> ^a	At1g34790	SALK_018552	T-DNA insertion or indel: T-DNA insertion or deletion in intron	Yellow seeds, brown spot at the chalazal-micropylar area	Crosses with <i>tt1-4</i> (SALK_107737)
<i>tt1-6</i> ^a	At1g34790	SALK_109926	Missense mutation: TGC to TGG; Cys To Trp at amino acid 198	Yellow seeds, brown spot at the chalazal-micropylar area	Crosses with <i>tt1-4</i> (SALK_107737)
<i>tt8-8</i> ^a	At4g09820	GK-089H05	Indel: 45 bp deletion in intron 6 of <i>TT8</i>	Yellow seeds	Crosses with <i>tt8-7</i> (SALK_082999)
<i>tt8-9</i> ^a	At4g09820	SALK_010879	Indel: 41 bp deletion including the end of intron 2 and the beginning of exon 3	Yellow seeds	Crosses with <i>tt8-7</i> (SALK_082999)
<i>tt8-10</i> ^a	At4g09820	SALK_137974	Unclear: a possible T-DNA insertion in intron 2 or a genome deletion including intron 2	Yellow seeds	Complementation with p35S::TT8 Crosses with <i>tt8-7</i> (SALK_082999)
<i>tt8-11</i> ^a	At4g09820	SALK_205661	Indel: 56 bp deletion at the end of intron 5 and the beginning of exon 6	Yellow seeds	Complementation with p35S::TT8 Crosses with <i>tt8-7</i> (SALK_082999)
<i>tt8-12</i> ^a	At4g09820	SALK_206210	T-DNA insertion: between 8 and 9 bp of intron 2	Yellow seeds	Crosses with <i>tt8-7</i> (SALK_082999)
<i>tt8-13</i> ^a	At4g09820	SAIL_239_C11	T-DNA insertion: between 127 and 128 bp of exon 2	Yellow seeds	Complementation with p35S::TT8 Crosses with <i>tt8-7</i> (SALK_082999) Complementation with p35S::TT8

* Indicates mutations confirmed by Sanger sequencing. Primer information for Sanger sequencing provided in Supplementary Table S7.
^a Indicates mutants identified by RNA-Seq.
^b Indicates mutants identified by whole genome re-sequencing.

lauroylsarcosine). Then, 0.7 mL mixture of phenol:chloroform:isoalcohol (25:24:1) was added and vortexed. After centrifugation, the upper aqueous phase (~0.7 mL) was mixed with 100 µL of 3 M NaOAc, and 1 mL of isopropanol for two hours at -20 °C to precipitate genomic DNAs. After incubation, the precipitated genomic DNAs were cleaned with 70 % ethanol and resuspended with water. 20 ng of gDNA was used as template in PCR.

The primers for confirming the indexed T-DNA insertions were designed by the iSect tool (<http://signal.salk.edu/tdnaprimers.2.html>) and synthesized by Integrated DNA Technologies (Skokie, IL, USA). PCR reactions were performed using Taq DNA Polymerases (NEB, MA, USA) followed by manufacturer protocols. PCR amplicons were then precipitated and purified for sequencing. The sequenced results were then aligned with *Arabidopsis* genomic sequence and located the indexed T-DNA insertion sites.

2.7. RNA-seq analysis

Total RNAs were extracted from *Arabidopsis* seedlings under AIC for four days using QIAGEN RNeasy Plant Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. Around 1 µg of RNA was used for construction of RNA-seq library (TruSeq Stranded mRNA, Illumina, CA, US). The resulting RNA-seq library with around 300 bp insertion sizes was sequenced with 50 single-end on an Illumina sequencing platform (Illumina HiSeq 4000, Ohio State University CCC-Genomics Shared Resource, OH, USA).

RNA-seq single-end reads were first evaluated for quality using FastQC (v.0.11.5) [65]. The reads were then aligned to the *Arabidopsis* reference genome (TAIR 10) using Hisat2 (v 2.1). Raw read counts for gene features was then quantified from these alignments using featureCounts (v1.6.2). Transcripts levels were then calculated in transcripts per million (TPM). Using the partial replicates of the library, R package DESeq2 [66] was employed to calculate the dispersion of the wild type samples and normalized log counts and to obtain the fold change of the 16 mutant samples.

2.8. Whole genome sequencing

Total genomic DNAs from pooled leaves of 30 *Arabidopsis* BC₁F₃ plants (*tt* phenotype) were extracted using Wizard Genomic DNA Purification Kit (Promega, WI, US). The pooled gDNA samples were then sent to BGI genomic services (<http://www.genomics.cn/en/index>) for 100 bp paired-end Illumina sequencing.

Quality of sequencing reads was evaluated using FastQC (v.0.11.5) [65] and a PHRED-scaled quality scores of 20 was set as the cutoff. In total, approximately 300 million paired-end reads were obtained (Supplementary Table S5). Reads were aligned using Bowtie2 (v 2.3.4) [67] to an *in silico* genome created by combining the reference TAIR10 *Arabidopsis thaliana* genome. A pseudo-chromosome was created by the flanking sequences of the T-DNA. After alignment screening of the mapped reads was performed to obtain junction reads which were defined as any read pairs where one mate mapped to one of the *Arabidopsis* chromosomes and the other to the T-DNA sequence (pseudo-chromosome). The junction reads were then annotated back to the corresponding genomic structure (gene or intergenic region). A visual confirmation of the identified sites was carried out by displaying on the Integrated genome browser (IGV).

2.9. Quantitative RT-PCR

Arabidopsis seedlings grown on MS media for seven days were used for RNA extraction using the QIAGEN RNeasy Plant Mini Kit (Qiagen, Germany). First-strand cDNA was synthesized from 1.5 µg of RNA using SuperScript II Reverse Transcriptase (ThermoFisher Scientific, CA, USA) and then diluted 1:5 with nuclease-free water. Quantitative PCR was performed in a 10 µL of mixture including 5 µL of SYBR Green PCR

Master Mix (ThermoFisher Scientific, CA, USA), 1.2 µL of gene specific primer pair (0.6 µL of each 5 µM forward and reverse primers), and 2 µL of the cDNA samples. Quantitative PCR was conducted by the ABI QuantStudio 7 Flex PCR system (ThermoFisher Scientific, CA, USA) with the following cycling conditions: 50 °C for 2 min, 95 °C for 10 min, then followed by 40 cycles of 95 °C for 15 s, 58 °C for 10 s, and 72 °C for 30 s. PCR results were processed and visualized using QuantStudio Real-Time PCR software v1.3 (ThermoFisher Scientific, CA, USA). The experiments were performed in four biological replicates and the housekeeping gene *Actin2* was used as the internal control for normalization of the expression levels. The $2^{-\Delta\Delta C_t}$ method was used to calculate the fold changes in gene expression among samples [68].

2.10. Large datasets

The raw reads of RNA-Seq and whole genome re-sequencing were deposited in the NCBI Gene Expression Omnibus and Sequence Read Archive under accession number GSE129589 and PRJNA532932, respectively.

3. Results

3.1. Characterization of transparent testa mutants from large T-DNA collections

A visual screening of seeds obtained from propagating at ABRC ~100,000 lines derived from the SALK, SAIL, and GABI indexed T-DNA collections resulted in the identification of 16 independent lines with seed coat colors different from the Col-0 wild type. The seed coat colors of the original seeds obtained from ABRC were often a mixture of normal dark brown and abnormal yellow or dark grey, suggesting that they had been collected from pools of segregating plants. Thus, seeds with the *tt* phenotypes from all the 16 lines were planted and plants allowed to self-pollinate to evaluate the uniformity of the color in the subsequent generation (M₁ seeds), and to further verify that no color segregation was present in the M₂ seeds, given that the seed coat is a maternal tissue (Table 1).

3.2. Indexed T-DNA insertions are not responsible for the transparent testa phenotypes

To determine which gene might be responsible for the mutant *tt* phenotype in each of the lines, we investigated in TAIR (<https://www.arabidopsis.org/>) the location of the respective indexed T-DNA insertions. We also confirmed by PCR that the F₁ plants corresponding to all 16 mutants were homozygous for the indexed T-DNA insertions (Supplementary Fig. S2). Suspiciously, none of the 16 genes or regions with the confirmed T-DNA insertions had been previously associated with flavonoid biosynthesis (Table 1 and Supplementary Table S2).

To determine whether the sequenced T-DNA insertions were indeed responsible for the observed *tt* phenotypes, we took a two-pronged approach (Fig. 1): For the first approach, when available, we obtained second mutant alleles corresponding to the indexed T-DNA insertion regions and evaluated the seed coat color in the self-pollinated plants (Supplementary Table S2). However, the seed coat colors from all the second alleles showed normal dark brown color, indistinguishable from Col-0, indicating that the indexed T-DNA insertion were unlikely responsible for the *tt* phenotypes.

For the second approach, we backcrossed each of the new mutants to Col-0, grew the BC₁F₁ seeds and allowed them to self-pollinate (BC₁F₂ seeds, Fig. 1). The BC₁F₂ plants were then analyzed by PCR for the presence of the indexed T-DNA as homozygous or heterozygous, and the seed coat colors of BC₁F₃ seeds were compared with the genotypes of the T-DNA insertion alleles. The *tt* phenotypes were observed from all seeds regardless of homo/heterozygosity for the indexed T-DNA insertion. In addition, the color of all BC₁F₁ seeds was colorless as

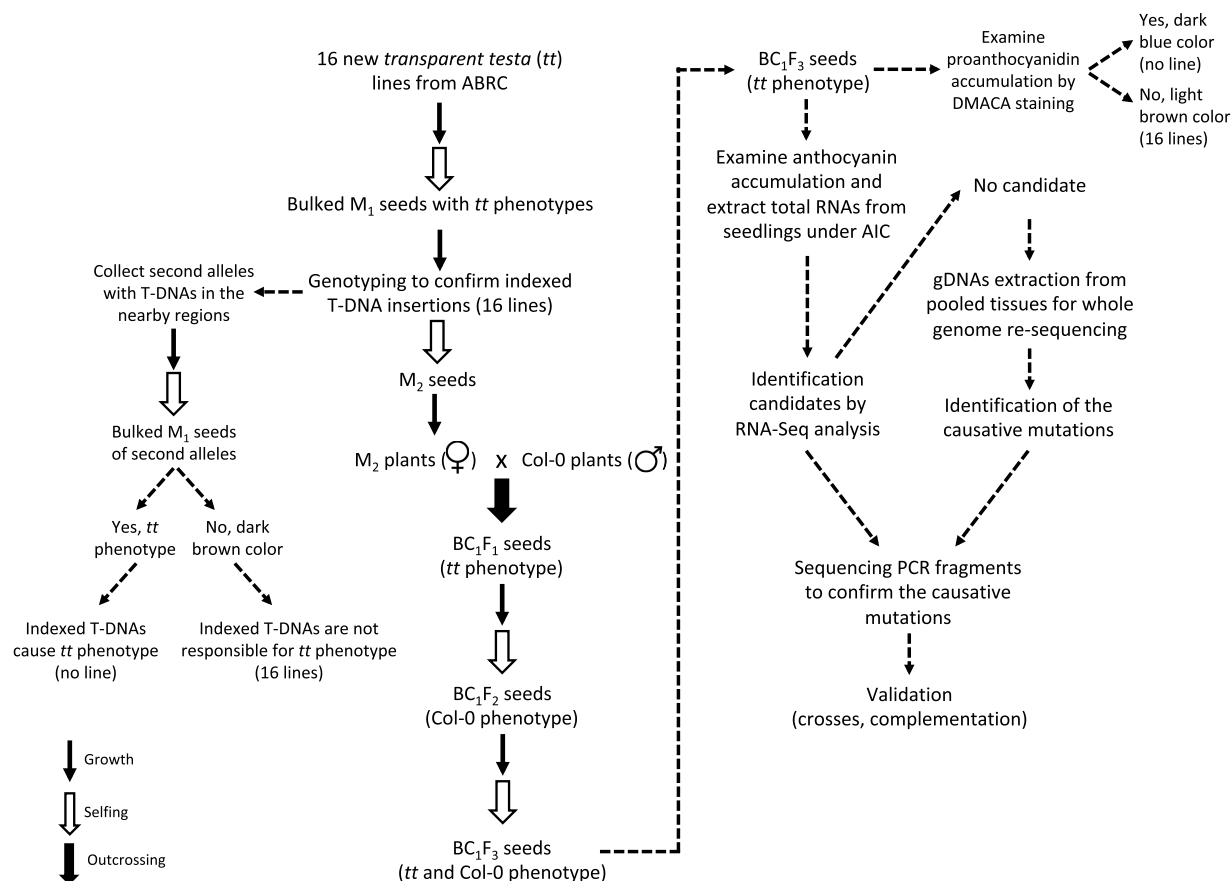


Fig. 1. Flowchart describing the step-by-step strategy to identify and validate the causative elements for the phenotypes in the 16 new *tt* mutants.

the same as in the *tt* maternal plants, while all BC₁F₂ seeds displayed brown wild-type color. In all BC₁F₃ seeds, Chi-square test validated that the segregation ratios between colorless and brown seeds were ~1:3 (Supplementary Table S3). These results indicate that additional monogenic recessive mutations elsewhere in the genome are responsible for the *tt* phenotypes in all the 16 newly identified mutant lines.

3.3. Phenotypes of new transparent testa alleles

To assess if the new *tt* mutant alleles affected solely PA formation or anthocyanin accumulation as well, PA accumulation in the seed coats of all the mutants (BC₁F₃, Fig. 1) was examined by DMACA staining (Fig. 2). Compared with the brown seed coat color of Col-0, two lines (SALK_052000 and SAIL_874_C01) showed grey seed coat color, while SALK_018552 and SALK_109926 seeds were colorless displaying a brown spot at the chalazal-micropylar area (Fig. 2, first column from the left), which was confirmed by the DMACA staining (Fig. 2, second column from the left). The rest of the mutant lines have colorless seed coats.

Next, we evaluated the anthocyanin accumulation in vegetative tissues and also trichome formation in BC₁F₃ mutant plants, since several PA/anthocyanin mutants are also defective in leaf hair formation [69]. A visual observation followed by absorbance of acidic methanol extracts at 535 nm (A_{535}) revealed that anthocyanin pigments were absent or barely detectable in AIC-grown seedlings from six mutants (SALK_012389, SAIL_1175_C01, SAIL_649_G09, SAIL_1180_D08, SAIL_641_E10, and SALK_023403; Fig. 2, third column from the left). The remaining ten mutants had similar or higher levels of anthocyanin pigments, when compared to Col-0, indicated by the percentages of anthocyanins relative to wild type Col-0 in analyses performed in

biological triplicates. In addition, two out of the six anthocyanin-deficient mutants (SALK_012389 and SAIL_1175_C01) also displayed trichome-less (glabrous) phenotypes (Supplementary Fig. S3).

A detailed analysis of the anthocyanin profiles resulting from the *tt* mutations was performed by HPLC on seedlings grown in AIC in the presence or absence of Nar (Fig. 2). AIC enables very young seedlings to accumulate high anthocyanin levels within a few days, while the addition of Nar permits to determine if the mutation causes a block in the pathway upstream or downstream of CHI (Supplementary Fig. S1) [70]. Thus, we investigated anthocyanin profiles in all the 16 mutants after feeding 100 μ M Nar for 24 h (AIC + Nar, Fig. 2). Four mutants (SALK_012389, SAIL_1175_C01, SAIL_649_G09, and SAIL_1180_D08) accumulated no anthocyanins, irrespective of whether Nar was supplied or not (top four mutant lines in Fig. 2), indicating that those lines harbored mutations in pathway genes downstream of CHI, or in pathway regulators. In contrast, Nar restored the pigmentation of SAIL_641_E10 and SALK_023403, indicating that these lines harbored mutations in *TT5*, *TT4* or a gene further upstream in the pathway. The remaining ten mutants showed normal anthocyanin pigmentation in AIC, albeit some quantitatively different from Col-0 (Fig. 2).

Based on all the observed phenotypes from the 16 mutants, we categorized them into six groups (Fig. 2): Group 1, two mutants (SALK_012389 and SAIL_1175_C01) with glabrous phenotypes and unable to accumulate PAs or anthocyanins. Group 2, two mutants (SAIL_649_G09 and SAIL_1180_D08) unable to accumulate PAs or anthocyanins, but displaying normal trichomes. Group 3, two mutants (SAIL_641_E10 and SALK_023403) unable to accumulate PAs or anthocyanins, but anthocyanin pigmentation is recovered following Nar addition. Group 4, two mutants (SALK_052000 and SAIL_874_C01) appear defective in the PA pathway and have dark grey seed coat color. Group 5, two mutants (SALK_018552 and SALK_109926) also defective

in the PA pathway except for PA accumulation at the chalazal-micropylar area. Both Group 5 mutants accumulate significantly higher amounts of anthocyanins than Col-0 under AIC. Group 6, six mutants (GK-089H05, SALK_010879, SALK_137974, SALK_205661, SALK_206210, and SAIL_239_C11) that fail to accumulate PAs, but have similar anthocyanin levels and profiles as Col-0.

3.4. RNA-seq and whole genome re-sequencing identify the loci responsible for the transparent testa phenotypes

As a strategy to characterize candidate loci defective in these 16

mutants, we conducted RNA-Seq analyses with mRNAs obtained from seedlings grown under AIC for four days to examine global changes of gene expression against Col-0, especially focusing on genes involved in flavonoid biosynthesis, regulation and sequestration (Supplementary Table S1). This analysis was performed considering the phenotypic characterization described in the previous section. Because the objective was to score presence/absence of pathway gene transcripts or abnormal read distribution (including abnormal transcript structures) using a genome browser, and in that way infer the provenance of the mutant phenotype, only one RNA-Seq replicate for each of the mutants and two replicates for Col-0 were initially done (Fig. 3 and

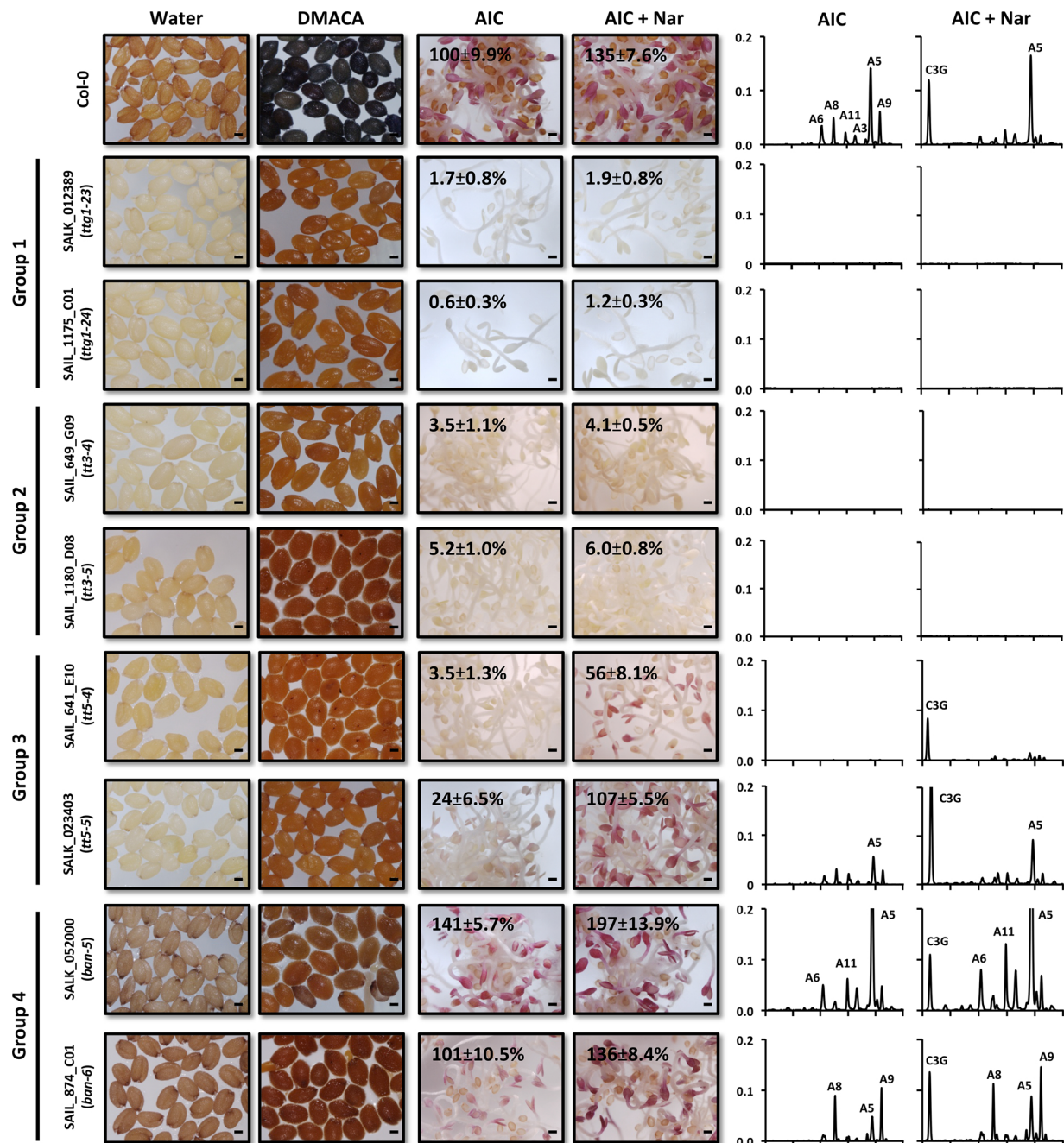


Fig. 2. Phenotype and characterization of 16 new *tt* mutants from the ABRC T-DNA collections. First column from the left, yellow and dark grey seed phenotypes. Second column, no or limited PA accumulation in seed coats with DMACA staining. Third to sixth columns, phenotypes and anthocyanin profiles from seedlings growth in AIC without and with 100 μ m of naringenin (Nar) treatment. Numbers in the third and fourth column represent the percentage (mean $\pm$ SD) of anthocyanin amount (evaluated by A_{535}) from triplicate extractions compared to Col-0 extraction under AIC as 100 %. C3G, cyanidin 3-glucoside; Six different anthocyanins (A3/A5/A6/A8/A9/A11) were characterized by LC-MS/MS [64]. Bar, 100 μ m. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

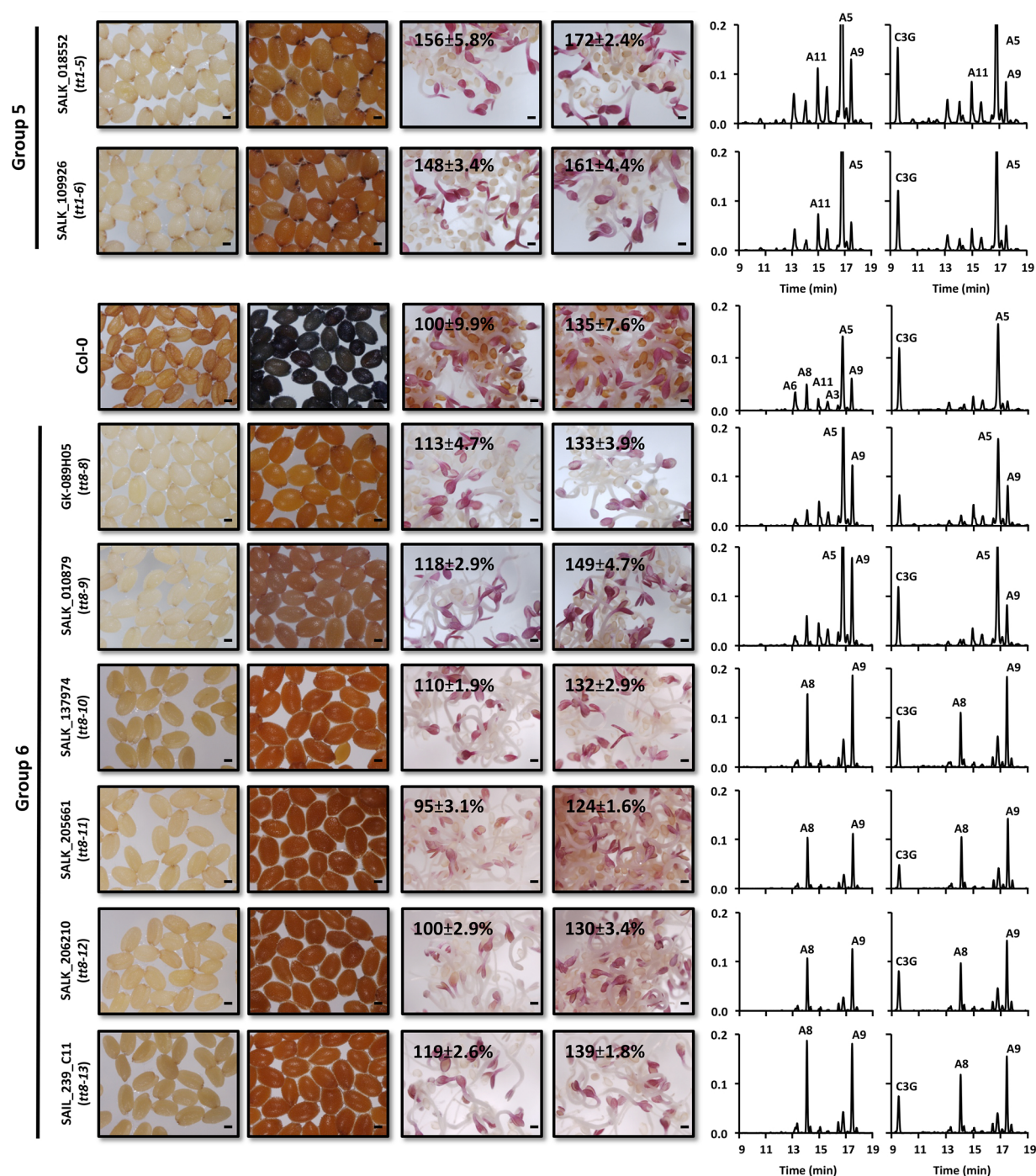


Fig. 2. (continued)

Supplementary Table S4). This analysis permitted us to hypothesize that two of the mutant alleles corresponded to *ttg1*, two to *tt3*, two to *tt5* and six to *tt8*. We provide below a brief description of how the RNA-Seq results suggested lesions in these genes.

RNA-Seq analysis revealed a large region in the exon of *TTG1* with no reads in SALK_012389 and complete absence of expression of *TTG1* and four nearby genes (At5g24490, At5g24500, At5g24510, and At5g24530) in SAIL_1175_C01 (Fig. 4A and B). In the SALK_012389 and SAIL_1175_C01 mutants, which we tentatively named *ttg1-23* and *ttg1-24* respectively, transcript abundances of many flavonoid pathway genes including core flavonoid genes (*TT4*, *TT5*, *TT6*, *TT7*, *TT3*, *LDOX*, and *TT19*), anthocyanin decoration genes (*3GT*, *UGT79B1*, *3AT1*, *3AT2*, *5GT*, *5MAT*, and *SAT*), and a regulatory gene (*TT8*) were significantly

reduced (Fig. 3). The reduced mRNA accumulation of anthocyanin biosynthesis genes in the mutants is consistent with previous results that showed *TTG1* controls several late anthocyanin pathway genes as well as *TT8* [41,51]. Phenotypically, the two mutants displayed the characteristic *ttg1* glabrous phenotypes (Table 1 and Supplementary Fig. S3), and no anthocyanin or PA production (Fig. 2).

Under AIC conditions (either with or without Nar), two mutants (SAIL_649_G09 and SAIL_1180_D08) failed to accumulate anthocyanins, nor did they accumulate seed coat PAs (Fig. 2). RNA-Seq analyses of these mutants showed minor effects on mRNA levels for other pathway genes (Fig. 3), and examination of the RNA-Seq reads on each of the pathway genes revealed that these two mutants had mutations in the *TT3* gene. In SAIL_649_G09, tentatively named *tt3-4*, a point mutation

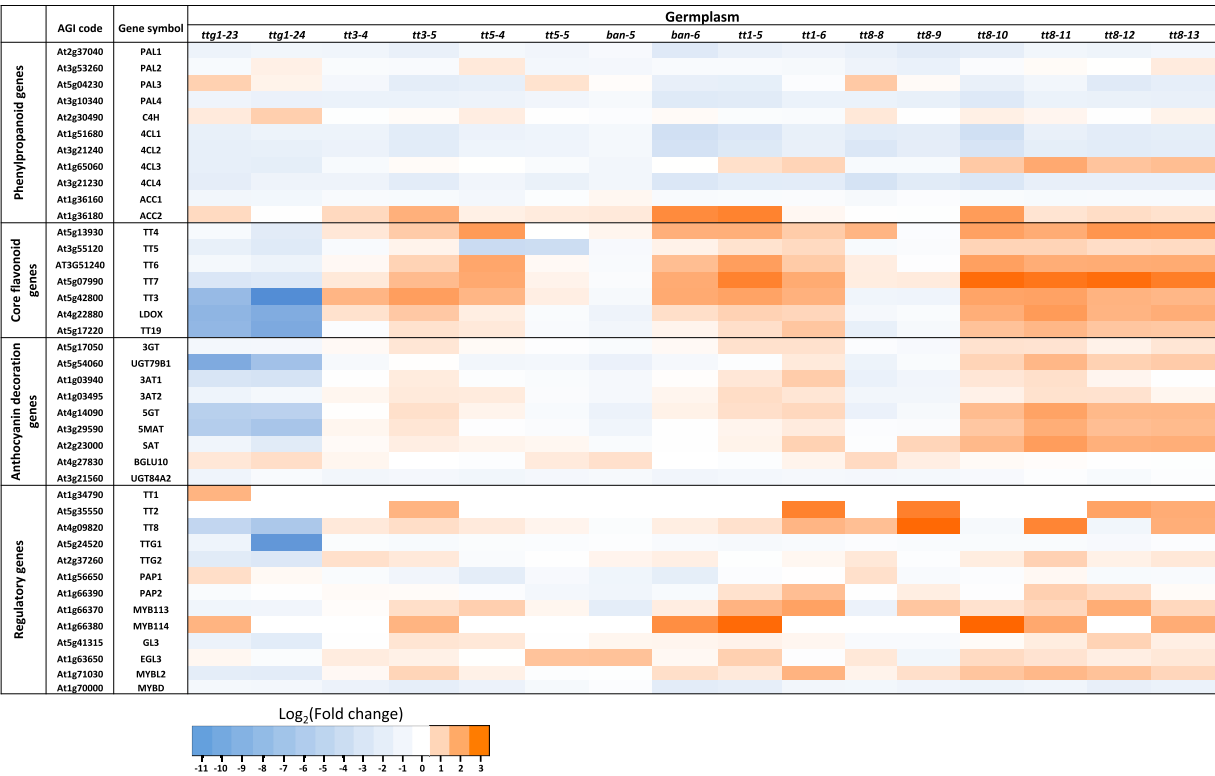


Fig. 3. Heat map summarizing gene transcript alterations for known loci related to flavonoid biosynthesis, regulation, and sequestration in 16 new *tt* mutants against Col-0 by RNA-Seq analysis. *ttg1-23*, SALK_012389; *ttg1-24*, SAIL_1175_C01; *tt3-4*, SAIL_649_G09; *tt3-5*, SAIL_1180_D08; *tt5-4*, SAIL_641_E10; *tt5-5*, SALK_023403; *ban-5*, SALK_052000; *ban-6*, SAIL_874_C01; *tt1-5*, SALK_018552; *tt1-6*, SALK_109926; *tt8-8*, GK-089H05; *tt8-9*, SALK_010879; *tt8-10*, SALK_137974; *tt8-11*, SALK_205661; *tt8-12*, SALK_206210; and *tt8-13*, SAIL_239_C11.

(G to A) was found from mapped reads in exon 3, which leads to a missense mutation (GGA to AGA; Gly to Arg) at position 130 (G130A) of the encoded DFR protein (Fig. 4C). In SAIL_1180_D08, tentatively named *tt3-5*, an indel (3 bp deletion, resulting in the deletion of D233) was identified in the fifth exon of *TT3* (Fig. 4C). The crystal structure of the DFR protein is available [71] and the G130A mutation is likely to affect the substrate binding pocket and substrate specificity (Supplementary Fig. S4 and S5) [72,92]. However, the structure does not help understand the phenotype of the D233 deletion in SAIL_1180_D08, although this residue is conserved in DFR proteins from several plants (Supplementary Fig. S4).

Based on the RNA-Seq profiles (Fig. 3) combined with the anthocyanin-deficient phenotype that was complemented by Nar (Fig. 2), we hypothesized that SAIL_641_E10 resulted from a T-DNA insertion or genome deletion in *TT5*. In the RNA-Seq results of SAIL_641_E10, no reads mapping to exons 2, 3, and 4 of *TT5* were found. Instead, a large number of reads were aligned to the intergenic downstream region of *TT5* (Fig. 4D, red trace). We named this allele *tt5-4*. The RNA-Seq results of SALK_023403 showed a significantly lower number of *TT5* reads, compared to Col-0 (Fig. 4D, blue trace), but we could not find any obvious mutation in the transcript reads. Thus, we subjected this mutant to whole genome re-sequencing, and the results showed a T-DNA insertion in the third intron of *TT5* (Fig. 4E), and we tentatively called this allele *tt5-5*. Genomic PCR using T-DNA and *TT5* specific primers validated the T-DNA insertion sites between 68 and 69 bp of intron 3 in *tt5-5*. Accumulation of anthocyanin in AIC was completely absent in *tt5-4*, while *tt5-5* produced ~25 % of wild type (Fig. 2), consistent with the lower levels of mRNA accumulation (Fig. 4D). In both *tt5-4* and *tt5-5*, *TT5* mRNA levels were decreased compared to WT, but those of other flavonoid pathway genes were only slightly altered (Fig. 3).

The distribution of RNA-Seq reads for *TT8* was found to be

abnormal in GK-089H05 (tentatively named *tt8-8*), SALK_010879 (tentatively named *tt8-9*), SALK_137974 (tentatively named *tt8-10*), SALK_205661 (tentatively named *tt8-11*), SALK_206210 (tentatively named *tt8-12*), and SAIL_239_C11 (tentatively named *tt8-13*). In *tt8-8*, reads were mapped to part of intron 6 to create an extended version of exon 7 (Fig. 5, red box I). To determine the nature of the mutation, DNA fragments were PCR amplified from both genomic DNA and cDNA, and subjected to sequencing. The sequenced results identified a 45 bp deletion in intron 6 of *TT8*, which makes the abnormal *TT8* transcript contain additional 39 nt. In *tt8-9*, a large number of reads aligning to intron 2 were observed (Fig. 5, red box II). Sequencing of this region amplified from both genomic DNA and cDNA revealed that transcription of intron 2 was the consequence of loss of the 3' splicing site by a 41 bp deletion at the end of intron 2 and the beginning of exon 3. In *tt8-10*, the RNA-Seq data suggested a possible T-DNA insertion in intron 2 or a large deletion downstream of the second intron, since the abnormal *TT8* transcript only contains exon 1 and an extended exon 2 (Fig. 5, red box III). However, no genomic DNA fragments could be amplified using a T-DNA end-specific and multiple *TT8* gene-specific primers, suggesting either a truncated T-DNA or a large deletion. In *tt8-11*, mapped RNA-Seq reads revealed four nucleotides missing (GTAA) at the beginning of exon 6, resulting in the loss of the correct 3' splicing site between intron 5 and exon 6 (Fig. 5, red boxes IV). Genomic DNA PCR amplification and sequencing showed a 56 bp deletion at the end of intron 5 and the beginning of exon 6. In *tt8-12*, the absence of reads mapping to exons 3–7 suggested the possibility of a deletion or a T-DNA insertion in intron 2 (Fig. 5, red box V). Sequencing results of PCR fragments obtained by using a T-DNA-end and *TT8* specific primer pair from genomic DNA revealed a T-DNA inserted between base pairs 8 and 9 of intron 2, resulting in a truncated *TT8* transcript including just exons 1 and 2. In *tt8-13*, a partial *TT8* transcript that lacks exon 1 and part of exon 2 was detected by RNA-Seq, again suggesting a possible T-

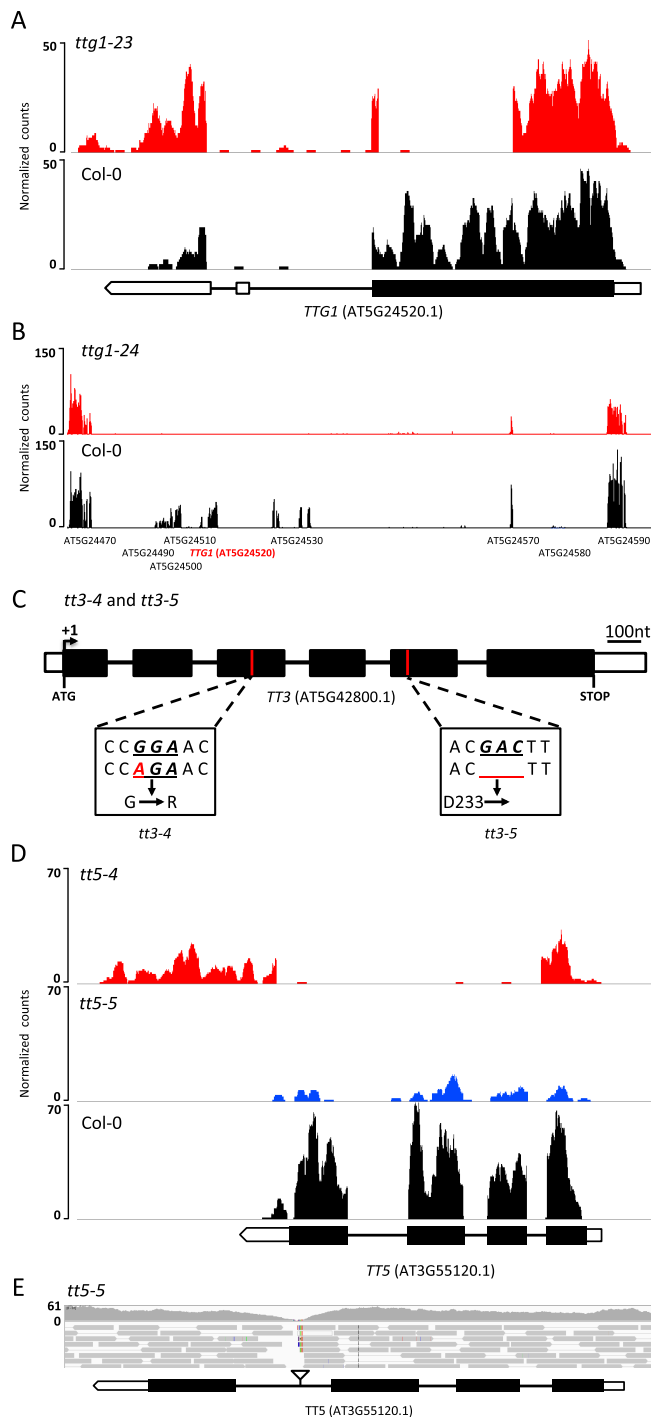


Fig. 4. Mapped results from RNA-Seq and whole genome re-sequencing analyses. (A) In *ttg1-23* (SALK_012389), a T-DNA insertion or deletion is present in exon 1 of *TTG1*; (B) *ttg1-24* (SAIL_1175_C01), a genome deletion across *TTG1*; (C) In *tt3-4* (SAIL_649_G09) and *tt3-5* (SAIL_1180_D08), a point mutation (G to A) and an indel (three bp deletion) were found in exons 3 and 5 of *TT3*, respectively; (D) In *tt5-4* (SAIL_641_E10), a truncated *TT5* is expressed with no exon 2–4, and lower mRNA accumulation of *TT5* in *tt5-5* (SALK_023403) was observed. (E) A T-DNA insertion in intron 3 of *TT5* was mapped by whole genome re-sequencing in *tt5-5*. The colored reads highlighted the mismatched sequence results to *Arabidopsis* genome that indicate candidate T-DNA insertion site. Exon and UTR are represented by closed and open boxes, respectively. Introns are indicated by a line between the exon boxes.

DNA insertion upstream of exon 2 (Fig. 5, red box VI). Indeed, a T-DNA insertion site was identified by PCR from genomic DNA between base pairs 127 and 128 of exon 2.

As expected from the function of *TT8* in regulating seed coat pigmentation [38,40], the six *tt8* mutant alleles showed significant reduction of PAs (Fig. 2), while the total amount of anthocyanin under AIC and the expression levels of anthocyanin biosynthesis genes in the mutant seedling were comparable to Col-0 (Figs. 2 and 3). We noticed that some seed batches of the *tt8-8* and *tt8-9* germinated at different rates, leading to anthocyanin (Fig. 2) and gene expression profiles that may be different from wild type, and also from the other four *tt8* alleles (Supplementary Table S5). The different anthocyanin profiles were in fact the reason why we subjected these mutants initially to RNA-Seq, rather than to whole genome re-sequencing. More experiments will need to be performed to establish to what extent these results suggest a minor participation of *TT8* in anthocyanin accumulation under AIC, or may merely reflect different anthocyanin and gene expression profiles as a consequence of seedling age heterogeneity because of germination initiation differences.

Group 5 mutants (SALK_018552 and SALK_109926) accumulate more anthocyanins than Col-0 under AIC (Fig. 2), but we could not infer from the expression results which gene(s) may harbor the mutation(s) (Fig. 3). Thus, we applied whole genome re-sequencing and identified these two mutants as new alleles of *TT1*, named *tt1-5* and *tt1-6*, respectively (Fig. 6A and B). A possible insertion was mapped to the intron of *TT1* in *tt1-5* (Fig. 6A), but genomic PCR using T-DNA and *TT1*-specific primers was not able to amplify a DNA fragment. A C/G point mutation that results in a missense mutation (C198W) was found in the aligned reads in *tt1-6* (Fig. 6B). The phenotypes of *tt1-5* and *tt1-6* are consistent with results that reported PAs accumulated only at the chalazal-micropylar area in *tt1* mutants (Fig. 2) [44,73].

Similar to Group 5 mutants, but distinct from SAIL_874_C01 (the other Group 4 mutant), SALK_052000 accumulated more anthocyanins in AIC than Col-0 (Fig. 2). Once again, we applied whole genome re-sequencing and determined that the two Group 4 mutants correspond to new alleles of *BAN*, tentatively named *ban-5* and *ban-6* (Fig. 6C and D). These new *ban* alleles have the same phenotypes as the known *BAN* mutant (*ban-4*, SALK_040250) that showed dark grey seed coat color (Fig. 2) [74]. In *ban-5*, we found a 27 bp in-frame deletion in exon 4 (Fig. 6C). The *ban-6* allele also showed a 10 bp deletion in exon 4 (Fig. 6D).

3.5. Validation of the identified loci by complementation

Since all 16 new mutants were preliminarily identified as new alleles of known flavonoid genes (two alleles of *TTG1*, two alleles of *TT3*, two alleles of *TT5*, two alleles of *BAN*, two alleles of *TT1*, and six alleles of *TT8*), two approaches were used to unequivocally verify the identity of the mutations (Table 1). The first method was to investigate segregation of the *tt* phenotypes by crossing our newly identified alleles with characterized (reference) alleles of the same flavonoid genes and then examining the seed coat color of the BC₁F₂ progenies. Six T-DNA insertion alleles including *ttg1-21* (GK-580A05), *tt3-2* (GK-295C10), *tt5-2* (GK-176H03), *ban-4* (SALK_040250), *tt1-4* (SALK_107737) and *tt8-7* (SALK_082999) were used as references for crossing the 16 new mutants [6,74,75]. The seeds of the F₂ progenies were analyzed, and the seed coat colors from all the crosses showed the *tt* phenotypes, without any color segregation (Supplementary Fig. S6). The second approach was to examine whether the *tt* phenotypes could be complemented by transformation with the corresponding functional genes. The *ttg1-23* and *ttg1-24* alleles were transformed with functional *TTG1* driven by the *CaMV 35S* promoter (p35S::TTG1) and the seed coat color of the T₁ seeds was investigated in at least two independent lines for each transformant. As observed in Supplementary Fig. S7, the seed coat phenotypes of *ttg1-23* and *ttg1-24* were restored to dark brown in each of the two lines investigated. In addition, the leaves of T₂ plants produced normal trichomes, complementing the glabrous phenotypes (Supplementary Fig. S3). Similarly, we transformed the four new *tt8* mutant alleles (*tt8-9*, *tt8-10*, *tt8-12*, and *tt8-13*) with p35S::TT8. The

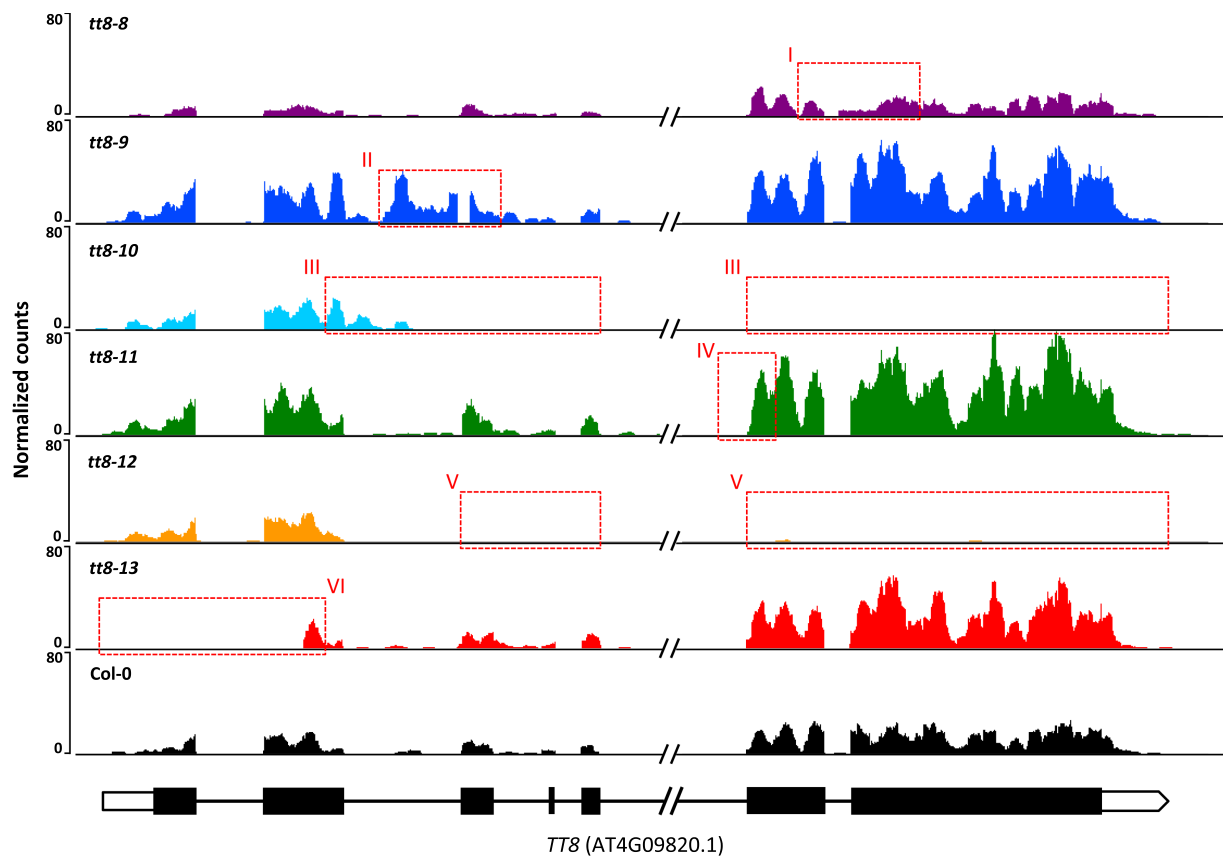


Fig. 5. Six abnormal *Tt8* transcripts were identified by RNA-Seq analysis. Extra reads mapped to part of intron 6 and intron 2 in *tt8-8* (GK-089H05, I) and *tt8-9* (SALK_010879, II), respectively. Three candidate T-DNA insertion sites were located in intron 2 of *tt8-10* (SALK_137974, III) and *tt8-12* (SALK_206210, V), and exon 2 of *tt8-13* (SAIL_239_C11, VI). In *tt8-11* (SALK_205661, IV), reads covered entire exon 6, except for the beginning four bp. Dashed red squares highlight the abnormal or missing reads. Exons and UTRs are represented by closed and open boxes, respectively. Introns are indicated by line between exons. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

transgene restored the brown seed coat colors of the *T₁* seeds from *tt8-9* and *tt8-13* (again, two independent lines for each shown) was observed (Supplementary Fig. S7). However, the same construct (p35S::Tt8) only partially restored the seed coat phenotype of *tt8-12* to light brown, and we observed no seed coat pigment accumulation for the transformed *tt8-10* allele in any of the 21 transgenic lines investigated (Supplementary Fig. S7). To determine the presence of functional *Tt8* transcripts in the transformed *tt8* lines, RT-qPCR experiments were performed using four primer pairs spanning across exons 1–2, exons 2–3, exons 5–6, and exons 6–7 (Supplementary Fig. S8). The RT-qPCR results confirmed the expression of full-length *Tt8* transcripts in the complemented transformed *tt8-10* and *tt8-12* lines. However, the expression levels of full-length *Tt8* mRNA levels were comparable in the transformed (but partially or non-complemented) *tt8-10* and *tt8-12* lines, with those in the complemented *tt8-9* and *tt8-13* lines (Supplementary Fig. S8). Thus, *Tt8* transgene expression is not the reason for partial complementation by p35S::Tt8 of the *tt8-10* and *tt8-12* mutants. However, the crosses of these alleles to *tt8-7* showed that *tt8-10* and *tt8-12* are *bona fide* *tt8* alleles (Supplementary Fig. S6), suggesting that some aspect of the mutant *tt8-9* and *tt8-13* alleles interferes with transgene function. Understanding this will require additional investigation.

4. Discussion

The screening of large collections of indexed insertional mutants for forward genetics has, and continues to be, a powerful tool in the quest to establish plant gene function [76,77]. Thanks to the collaborative nature of many members of the *Arabidopsis* community, the T-DNA

insertion mutant collection has grown over the past couple of decades to over 325,000 lines [59,78–83]. Secondary mutations have been identified in large-scale genetic screenings using the T-DNA collections, although they have been rather infrequent [84,85]. From this perspective, our results are surprising in that all the 16 *tt* mutants phenotypically identified fail to correspond to the indexed T-DNA insertions. Additional T-DNA insertions are frequent and the average number of insertions per line in the SALK/GABI T-DNA collection was estimated to be at least 1.5 [86]. Thus, it might have been expected that several of the *tt* phenotypes were caused by a second insertion. However, out of the 16 new alleles, only three had obvious second T-DNA insertions (one in intron 3 of *Tt5*, one in exon 2 of *Tt8*, and one in intron 2 of *Tt8*). The other mutations included eight deletions (one in the exon of *TtG1*, one in exon 5 of *Tt3*, one from intron 1 to exon 4 of *Tt5*, two in exon 4 of *BAN*, and one each in introns 2, 5, and 6 of *Tt8*), two missense mutations (in *Tt3* and *Tt1*), one large deletion in *TtG1*, and two insertion/deletion of unclear origin in *Tt1* and *Tt8*. This high frequency of secondary mutations not associated with intact T-DNA elements is something that needs to be considered when using the T-DNA insertional collections for forward genetics screens. The identified mutations could correspond to spontaneous mutations that arose during the propagation of the stocks [85], or from unsuccessful T-DNA insertions, which can result in small insertion/deletions [93].

Identifying secondary mutations responsible for a phenotype can be challenging. We took advantage of the pigment distribution in seedlings under AIC, with and without Nar, and in the seed coat to classify the mutants into six groups. We then used a ‘quick and dirty’ single RNA-Seq replicate obtained from seedlings grown in AIC (and therefore expressing all the pathway genes at high levels) to determine if many

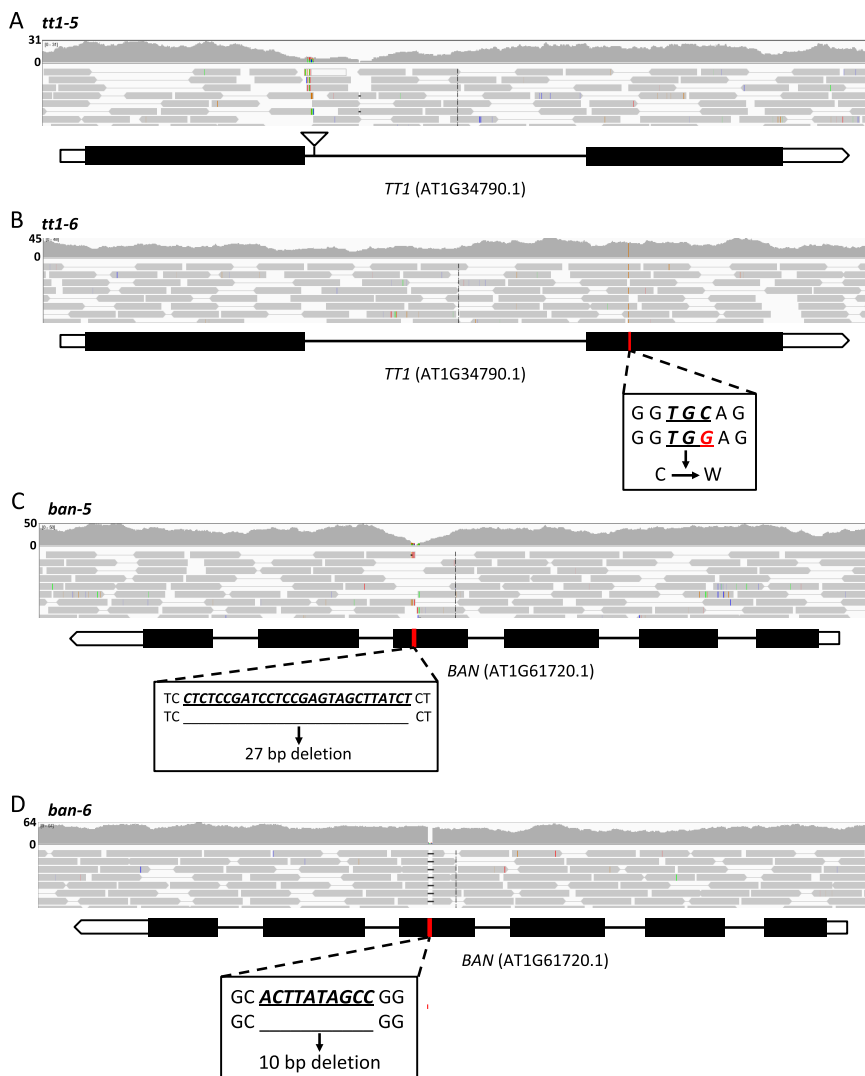


Fig. 6. Whole genome re-sequencing helps identify causative mutations for the *tt* phenotypes. In *tt1-5* (SALK_018552), a candidate T-DNA insertion is found in intron 1 of *TT1* (A). In *tt1-6* (SALK_109926), a point mutation (C to G) is found in exon 2 that creates a missense mutation in *TT1* (B). Two indels (27 bp and 10 bp deletions) in exon 4 of *BAN* were identified in *ban-5* (SALK_052000, C) and *ban-6* (SAIL_874_C01, D), respectively. The colored reads highlighted the mismatched sequence results to *Arabidopsis* genome that indicate candidate T-DNA insertion site or genetic lesion. Exons and UTRs are represented by closed and open boxes, respectively. Introns are indicated by lines between the exons.

pathway genes were coordinately affected (indicative of a regulatory mutation), and to align the reads against a set of 83 genes that we identified as necessary for pigment formation (Supplementary Table S5). Only in those instances in which we could not find a candidate gene based on the metabolic or RNA-Seq analyses, we appealed to whole genome re-sequencing (Supplementary Table S6). This sequential combination of approaches resulted in the relatively fast and inexpensive identification of the gene defects responsible for the 16 mutations, which were ultimately validated by complementation and crosses.

One unexpected result was the significantly higher anthocyanin accumulation levels in seedlings under AIC for both *tt1-5* and *tt1-6* mutants (Fig. 2). TT1/WIP1 belongs to a six-member WIP domain protein family (WIP1-6) or named the A1d subgroup in the A1 family of *Arabidopsis* C2H2 zinc finger proteins (AT-ZFPs). TT1 and the other members of the A1 family contain four zinc finger domains (ZF1-4) [87]. In the *tt1-6* mutant, the seed coat phenotype and higher anthocyanin accumulation are likely caused by the C198W mutation. Previous studies identified C198 as the second cysteine in the “CX4CX20HX4H” pattern of ZF2 that is predicted as a critical zinc coordinating residue [88]. TT1 has been characterized as a transcriptional regulator that is involved in controlling PA accumulation in the seed coat endothelium, with no obvious effects on anthocyanin biosynthesis [44,45]. Moreover, other WIP proteins (WIP2 and WIP6) from the A1d subgroup are involved in developmental processes [89,90]. Thus, TT1

could participate in controlling seedling development, and the observed increase in anthocyanin accumulation is an indirect consequence of this. Related to this, previous studies in *Solanum tuberosum* determined that a member of the A1a subgroup, StPCP1, acts as a transcriptional activator of a sucrose uptake transporter [91]. From this evidence, the function of TT1 in a high sucrose environment could be to act as a transcriptional repressor of sucrose transport. Thus, in *tt1* mutants, more sucrose would be taken-up by the seedlings, resulting in increased anthocyanin accumulation.

In conclusion, our unbiased analysis of indexed *Arabidopsis* T-DNA insertion lines resulted in the identification of 16 new mutants affecting flavonoid pigment formation. However, none of the indexed insertions is responsible for the mutants ensuing the *tt* phenotypes. This result should be a call to the community to be cautious when interpreting the results of forward-genetic screens using T-DNA insertion collections, and highlight the high frequency of mutations present in these lines. The identified alleles also provide interesting new information regarding the role of PA biosynthesis enzymes and regulators on anthocyanin accumulation.

5. Author contributions

LR identified the mutant lines; NJ carried out the characterization of the mutants and the phenotypes; NJ performed complementation with the assistance of YSL; EM and FGC carried out the computational

analyses and data was interpreted by NJ and YSL; NJ, YSL, and EG wrote the article with contributions from all the authors; EG coordinated and supervised the research project and agrees to serve as the author responsible for contact and ensures communication.

Funding information

This study was partially funded by grant NSFMCB-1822343 to EG and by the NSF Research Traineeship Program (DGE-1828149) to FGC.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgment

We thank members of the ABRC for their assistance with identifying lines with seed coat pigmentation defects and for their continuous support of the community over the years.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.plantsci.2019.110335>.

References

- [1] B. Winkel-Shirley, It takes a garden. How work on diverse plant species has contributed to an understanding of flavonoid metabolism, *Plant Physiol.* 127 (2001) 1399–1404.
- [2] R. Koes, W. Verweij, F. Quattrocchio, Flavonoids: a colorful model for the regulation and evolution of biochemical pathways, *Trends Plant Sci.* 10 (2005) 236–242.
- [3] L. Lepiniec, I. Debeaujon, J.M. Routaboul, A. Baudry, L. Pourcel, N. Nesi, M. Caboche, Genetics and biochemistry of seed flavonoids, *Annu. Rev. Plant Biol.* 57 (2006) 405–430.
- [4] M. Koornneef, Mutations affecting the testa color in *Arabidopsis*, *Arabidopsis Inf. Serv.* 28 (1990) 1–4.
- [5] B.W. Shirley, W.L. Kubasek, G. Storz, E. Bruggemann, M. Koornneef, F.M. Ausubel, H.M. Goodman, Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis, *Plant J.* 8 (1995) 659–671.
- [6] I. Appelhagen, K. Thiedig, N. Nordholt, N. Schmidt, G. Huep, M. Sagasser, B. Weisshaar, Update on *transparent testa* mutants from *Arabidopsis thaliana*: characterisation of new alleles from an isogenic collection, *Planta* 240 (2014) 955–970.
- [7] D.E. Saslow, C.D. Dana, B. Winkel-Shirley, An allelic series for the chalcone synthase locus in *Arabidopsis*, *Gene* 255 (2000) 127–138.
- [8] B.W. Shirley, S. Hanley, H.M. Goodman, Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis transparent testa* mutations, *Plant Cell* 4 (1992) 333–347.
- [9] M.K. Pelletier, B.W. Shirley, Analysis of flavanone 3-hydroxylase in *Arabidopsis* seedlings. Coordinate regulation with chalcone synthase and chalcone isomerase, *Plant Physiol.* 111 (1996) 339–345.
- [10] E. Wisman, U. Hartmann, M. Sagasser, E. Baumann, K. Palme, K. Hahlbrock, H. Saedler, B. Weisshaar, Knock-out mutants from an *En-1* mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes, *Proc. Natl. Acad. Sci. U. S. A.* 95 (1998) 12432–12437.
- [11] C. Schoenbohm, S. Martens, C. Eder, G. Forkmann, B. Weisshaar, Identification of the *Arabidopsis thaliana* flavonoid 3'-hydroxylase gene and functional expression of the encoded P450 enzyme, *Biol. Chem.* 381 (2000) 749–753.
- [12] M.K. Pelletier, J.R. Murrell, B.W. Shirley, Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in *Arabidopsis*. Further evidence for differential regulation of "early" and "late" genes, *Plant Physiol.* 113 (1997) 1437–1445.
- [13] A. Preuss, R. Stracke, B. Weisshaar, A. Hillebrecht, U. Matern, S. Martens, *Arabidopsis thaliana* expresses a second functional flavonol synthase, *FEBS Lett.* 583 (2009) 1981–1986.
- [14] E.T. Johnson, S. Ryu, H. Yi, B. Shin, H. Cheong, G. Choi, Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase, *Plant J.* 25 (2001) 325–333.
- [15] J. Nakajima, Y. Tanaka, M. Yamazaki, K. Saito, Reaction mechanism from leucoanthocyanidin to anthocyanidin 3-glucoside, a key reaction for coloring in anthocyanin biosynthesis, *J. Biol. Chem.* 276 (2001) 25797–25803.
- [16] S. Abrahams, E. Lee, A.R. Walker, G.J. Tanner, P.J. Larkin, A.R. Ashton, The *Arabidopsis TDS4* gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development, *Plant J.* 35 (2003) 624–636.
- [17] I. Appelhagen, O. Jahns, L. Bartelniewoehner, M. Sagasser, B. Weisshaar, R. Stracke, Leucoanthocyanidin Dioxygenase in *Arabidopsis thaliana*: characterization of mutant alleles and regulation by MYB-BHLH-TTG1 transcription factor complexes, *Gene* 484 (2011) 61–68.
- [18] S. Albert, M. Delseny, M. Devic, *BANYULS*, a novel negative regulator of flavonoid biosynthesis in the *Arabidopsis* seed coat, *Plant J.* 11 (1997) 289–299.
- [19] M. Devic, J. Guilleminot, I. Debeaujon, N. Bechtold, E. Bensaude, M. Koornneef, G. Pelletier, M. Delseny, The *BANYULS* gene encodes a DFR-like protein and is a marker of early seed coat development, *Plant J.* 19 (1999) 387–398.
- [20] D.Y. Xie, S.B. Sharma, N.L. Paiva, D. Ferreira, R.A. Dixon, Role of anthocyanidin reductase, encoded by *BANYULS* in plant flavonoid biosynthesis, *Science* 299 (2003) 396–399.
- [21] T. Ichino, K. Fuji, H. Ueda, H. Takahashi, Y. Koumoto, J. Takagi, K. Tamura, R. Sasaki, K. Aoki, T. Shimada, I. Hara-Nishimura, GFS9/TT9 contributes to intracellular membrane trafficking and flavonoid accumulation in *Arabidopsis thaliana*, *Plant J.* 80 (2014) 410–423.
- [22] S. Kitamura, N. Shikazono, A. Tanaka, *TRANSPARENT TESTA 19* is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*, *Plant J.* 37 (2004) 104–114.
- [23] I. Debeaujon, A.J. Peeters, K.M. Léon-Kloosterziel, M. Koornneef, The *TRANSPARENT TESTA12* gene of *Arabidopsis* encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium, *Plant Cell* 13 (2001) 853–871.
- [24] K. Marinova, L. Pourcel, B. Weder, M. Schwarz, D. Barron, J.M. Routaboul, I. Debeaujon, M. Klein, The *Arabidopsis* MATE transporter TT12 acts as a vacuolar flavonoid/H⁺-antiporter active in proanthocyanidin-accumulating cells of the seed coat, *Plant Cell* 19 (2007) 2023–2038.
- [25] I.R. Baxter, J.C. Young, G. Armstrong, N. Foster, N. Bogenschutz, T. Cordova, W.A. Peer, S.P. Hazen, A.S. Murphy, J.F. Harper, A plasma membrane H⁺-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 2649–2654.
- [26] I. Appelhagen, N. Nordholt, T. Seidel, K. Spelt, R. Koes, F. Quattrocchio, M. Sagasser, B. Weisshaar, *TRANSPARENT TESTA 13* is a tonoplast P_{3A}-ATPase required for vacuolar deposition of proanthocyanidins in *Arabidopsis thaliana* seeds, *Plant J.* 82 (2015) 840–849.
- [27] L. Pourcel, J.M. Routaboul, L. Kerhoas, M. Caboche, L. Lepiniec, I. Debeaujon, *TRANSPARENT TESTA10* encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in *Arabidopsis* seed coat, *Plant Cell* 17 (2005) 2966–2980.
- [28] T. Tohge, Y. Nishiyama, M.Y. Hirai, M. Yano, J. Nakajima, M. Awazu, E. Inoue, H. Takahashi, D.B. Goodenow, M. Kitayama, M. Noji, M. Yamazaki, K. Saito, Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor, *Plant J.* 42 (2005) 218–235.
- [29] K. Saito, K. Yonekura-Sakakibara, R. Nakabayashi, Y. Higashi, M. Yamazaki, T. Tohge, A.R. Fernie, The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity, *Plant Physiol. Biochem.* 72 (2013) 21–34.
- [30] N. Kovinch, G. Kayanja, A. Chanoca, K. Riedl, M.S. Otegui, E. Grotewold, Not all anthocyanins are born equal: distinct patterns induced by stress in *Arabidopsis*, *Planta* 240 (2014) 931–940.
- [31] N. Sasaki, Y. Nishizaki, Y. Ozeki, T. Miyahara, The role of acyl-glucose in anthocyanin modifications, *Molecules* 19 (2014) 18747–18766.
- [32] W. Xu, C. Dubos, L. Lepiniec, Transcriptional control of flavonoid biosynthesis by MYB-BHLH-WDR complexes, *Trends Plant Sci.* 20 (2015) 176–185.
- [33] A. Feller, K. Machemer, E.L. Braun, E. Grotewold, Evolutionary and comparative analysis of MYB and BHLH plant transcription factors, *Plant J.* 66 (2011) 94–116.
- [34] J. Liu, A. Osbourn, P. Ma, MYB transcription factors as regulators of phenylpropanoid metabolism in plants, *Mol. Plant* 8 (2015) 689–708.
- [35] J.C. Miller, W.R. Chezem, N.K. Clay, Ternary WD40 repeat-containing protein complexes: evolution, composition and roles in plant immunity, *Front. Plant Sci.* 6 (2016) 1108.
- [36] A. Lloyd, A. Brockman, L. Aguirre, A. Campbell, A. Bean, A. Cantero, A. Gonzalez, Advances in the MYB-BHLH-WD Repeat (MBW) Pigment regulatory model: addition of a WRKY factor and co-option of an anthocyanin MYB for betalain regulation, *Plant Cell Physiol.* 58 (2017) 1431–1441.
- [37] N. Nesi, C. Jond, I. Debeaujon, M. Caboche, L. Lepiniec, The *Arabidopsis* TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed, *Plant Cell* 13 (2001) 2099–2114.
- [38] N. Nesi, I. Debeaujon, C. Jond, G. Pelletier, M. Caboche, L. Lepiniec, The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques, *Plant Cell* 12 (2000) 1863–1878.
- [39] A.R. Walker, P.A. Davison, A.C. Bolognesi-Winfield, C.M. James, N. Srinivasan, T.L. Blundell, J.J. Esch, M.D. Marks, J.C. Gray, The *TRANSPARENT TESTA GLABRA1* locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein, *Plant Cell* 11 (1999) 1337–1350.
- [40] A. Baudry, M. Heim, B. Dubreucq, M. Caboche, B. Weisshaar, L. Lepiniec, TT2, TT8 and TTG1 synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *Arabidopsis thaliana*, *Plant J.* 39 (2004) 366–380.
- [41] A. Baudry, M. Caboche, L. Lepiniec, TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and BHLH factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*, *Plant J.* 46 (2006) 768–779.
- [42] C.S. Johnson, B. Koleski, D.R. Smyth, *TRANSPARENT TESTA GLABRA2*, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor, *Plant Cell* 14 (2002) 1359–1375.
- [43] A. Gonzalez, M. Brown, G. Hatlestad, N. Akhavan, T. Smith, A. Hemdb, J. Moore, D. Montes, T. Mosley, J. Resendez, H. Nguyen, L. Wilson, A. Campbell, D. Sudarshan, A. Lloyd, TTG2 controls the developmental regulation of seed coat

- tannins in *Arabidopsis* by regulating vacuolar transport steps in the proanthocyanidin pathway. *Dev. Biol.* 419 (2016) 54–63.
- [44] M. Sagasser, G.H. Lu, K. Hahlbrock, B. Weisshaar, A. *thaliana* TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins, *Genes Dev.* 16 (2002) 138–149.
- [45] I. Appelhagen, G.H. Lu, G. Huep, E. Schmelzer, B. Weisshaar, M. Sagasser, TRANSPARENT TESTA1 interacts with R2R3-MYB factors and affects early and late steps of flavonoid biosynthesis in the endothelium of *Arabidopsis thaliana* seeds, *Plant J.* 67 (2011) 406–419.
- [46] N. Focks, M. Sagasser, B. Weisshaar, C. Benning, Characterization of *tt15*, a novel transparent testa mutant of *Arabidopsis thaliana* (L.) Heynh, *Planta* 208 (1999) 352–357.
- [47] W. Xu, S. Bobet, J. Le Gourrierc, D. Grain, D. De Vos, A. Berger, F. Salsac, Z. Kelemen, J. Boucherez, A. Rolland, G. Mouille, J.M. Routaboul, L. Lepiniec, C. Dubos, TRANSPARENT TESTA 16 and 15 act through different mechanisms to control proanthocyanidin accumulation in *Arabidopsis* testa, *J. Exp. Bot.* 68 (2017) 2859–2870.
- [48] N. Nesi, I. Debeaujon, C. Jond, A.J. Stewart, G.I. Jenkins, M. Caboche, L. Lepiniec, The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat, *Plant Cell* 14 (2002) 2463–2479.
- [49] S. DeBolt, W.R. Scheible, K. Schrick, M. Auer, F. Beisson, V. Bischoff, P. Bouvier-Nave, A. Carroll, K. Hematy, Y. Li, J. Milne, M. Nair, H. Schaller, M. Zemla, C. Somerville, Mutations in UDP-glucose:sterol glucosyltransferase in *Arabidopsis* cause transparent testa phenotype and suberization defect in seeds, *Plant Physiol.* 151 (2009) 78–87.
- [50] J.O. Borevitz, Y. Xia, J. Blount, R.A. Dixon, C. Lamb, Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis, *Plant Cell* 12 (2000) 2383–2394.
- [51] A. Gonzalez, M. Zhao, J.M. Leavitt, A.M. Lloyd, Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings, *Plant J.* 53 (2008) 814–827.
- [52] C.T. Payne, F. Zhang, A.M. Lloyd, GL3 encodes a bHLH protein that regulates trichome development in *Arabidopsis* through interaction with GL1 and TTG1, *Genetics* 156 (2000) 1349–1362.
- [53] F. Zhang, A. Gonzalez, M. Zhao, C.T. Payne, A. Lloyd, A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*, *Development* 130 (2003) 4859–4869.
- [54] L. Meng, M. Xu, W. Wan, F. Yu, C. Li, J. Wang, Z. Wei, M. Lv, X. Cao, Z. Li, J. Jiang, Sucrose signaling regulates anthocyanin biosynthesis through a MAPK cascade in *Arabidopsis thaliana*, *Genetics* 210 (2018) 607.
- [55] B. Zhang, A. Schrader, TRANSPARENT TESTA GLABRA 1-dependent regulation of flavonoid biosynthesis, *Plants (Basel)* 6 (2017) 4.
- [56] C. Dubos, J. Le Gourrierc, A. Baudry, G. Huep, E. Lanet, I. Debeaujon, J.M. Routaboul, A. Alboresi, B. Weisshaar, L. Lepiniec, MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*, *Plant J.* 55 (2008) 940–953.
- [57] N.H. Nguyen, C.Y. Jeong, G.H. Kang, S.D. Yoo, S.W. Hong, H. Lee, MYBD employed by HY5 increases anthocyanin accumulation via repression of MYBL2 in *Arabidopsis*, *Plant J.* 84 (2015) 1192–1205.
- [58] H. Kubo, A.J. Peeters, M.G. Aarts, A. Pereira, M. Koornneef, ANTHOCYANINLESS2, a homeobox gene affecting anthocyanin distribution and root development in *Arabidopsis*, *Plant Cell* 11 (1999) 1217–1226.
- [59] J.M. Alonso, A.N. Stepanova, T.J. Leisse, C.J. Kim, H. Chen, P. Shinn, D.K. Stevenson, J. Zimmerman, P. Barajas, R. Cheuk, Genome-wide insertional mutagenesis of *Arabidopsis thaliana*, *Science* 301 (2003) 653–657.
- [60] T. Nakagawa, T. Suzuki, S. Murata, S. Nakamura, T. Hino, K. Maeo, R. Tabata, T. Kawai, K. Tanaka, Y. Niwa, Y. Watanabe, K. Nakamura, T. Kimura, S. Ishiguro, Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants, *Biosci. Biotechnol. Biochem.* 71 (2007) 2095–2100.
- [61] S.J. Clough, A.F. Bent, Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*, *Plant J.* 16 (1998) 735–743.
- [62] T. Murashige, F. Skoog, A revised medium for rapid growth and bioassays with tobacco tissue cultures, *Physiol. Plant.* 15 (1962) 473–497.
- [63] S. Abrahams, G.J. Tanner, P.J. Larkin, A.R. Ashton, Identification and biochemical characterization of mutants in the proanthocyanidin pathway in *Arabidopsis*, *Plant Physiol.* 130 (2002) 561–576.
- [64] L. Pourcel, N.G. Irani, Y. Lu, K. Riedl, S. Schwartz, E. Grotewold, The formation of anthocyanic vacuolar inclusions in *Arabidopsis thaliana* and implications for the sequestration of anthocyanin pigments, *Mol. Plant* 3 (2010) 78–90.
- [65] S. Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data, Available at: (2010) <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- [66] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.* 15 (2014) 550.
- [67] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie2, *Nat. Methods* 9 (2012) 357–359.
- [68] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method, *Methods* 25 (2001) 402–408.
- [69] S. Pattanaik, B. Patra, S.K. Singh, L. Yuan, An overview of the gene regulatory network controlling trichome development in the model plant, *Arabidopsis*, *Front. Plant Sci.* 5 (2014) 259.
- [70] F. Poustka, N.G. Irani, A. Feller, Y. Lu, L. Pourcel, K. Frame, E. Grotewold, A trafficking pathway for anthocyanins overlaps with the endoplasmic reticulum-to-vacuole protein sorting route in *Arabidopsis* and contributes to the formation of vacuolar inclusions, *Plant Physiol.* 145 (2007) 1323–1335.
- [71] M. Beld, C. Martin, H. Huits, A.R. Stuitje, A.G. Gerats, Flavonoid synthesis in *Petunia hybrida*: partial characterization of dihydroflavonol-4-reductase genes, *Plant Mol. Biol.* 13 (1989) 491–502.
- [72] P. Petit, T. Granier, B.L. d'Estaintot, C. Manigand, K. Bathany, J.M. Schmitter, V. Lauvergeat, S. Hamdi, B. Gallois, Crystal structure of grape dihydroflavonol 4-reductase, a key enzyme in flavonoid biosynthesis, *J. Mol. Biol.* 368 (2007) 1345–1357.
- [73] I. Debeaujon, N. Nesi, P. Perez, M. Devic, O. Grandjean, M. Caboche, L. Lepiniec, Proanthocyanidin-accumulating cells in *Arabidopsis* testa: regulation of differentiation and role in seed development, *Plant Cell* 11 (2003) 2514–2531.
- [74] P.A. Bowerman, M.V. Ramirez, M.B. Price, R.F. Helm, B.S. Winkel, Analysis of T-DNA alleles of flavonoid biosynthesis genes in *Arabidopsis* ecotype Columbia, *BMC Res. Notes* 5 (2012) 485.
- [75] T. Qi, S. Song, Q. Ren, D. Wu, H. Huang, Y. Chen, M. Fan, W. Peng, C. Ren, D. Xie, The Jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate Jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*, *Plant Cell* 23 (2011) 1795–1814.
- [76] J.M. Alonso, J.R. Ecker, Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*, *Nat. Rev. Genet.* 7 (2006) 524–536.
- [77] N.J. Nannas, R.K. Dawe, Genetic and genomic toolbox of *Zea mays*, *Genetics* 199 (2015) 655–669.
- [78] P.J. Krysan, J.C. Young, M.R. Sussman, T-DNA as an insertional mutagen in *Arabidopsis*, *Plant Cell* 11 (1999) 2283–2290.
- [79] M. Galbiati, M.A. Moreno, G. Nadzan, M. Zourelidou, S.L. Dellaporta, Large-scale T-DNA mutagenesis in *Arabidopsis* for functional genomic analysis, *Funct. Integr. Genomics* 1 (2000) 25–34.
- [80] F. Samson, V. Brunaud, S. Balzergue, B. Dubreucq, L. Lepiniec, G. Pelletier, M. Caboche, A. Lecharny, FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants, *Nucleic Acids Res.* 30 (2002) 94–97.
- [81] M.G. Rosso, Y. Li, N. Strizhov, B. Reiss, K. Dekker, B. Weisshaar, An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics, *Plant Mol. Biol.* 53 (2003) 247–259.
- [82] R.C. O'Malley, J.R. Ecker, Linking genotype to phenotype using the *Arabidopsis* unimutant collection, *Plant J.* 61 (2010) 928–940.
- [83] N.J. Provart, J. Alonso, S.M. Assmann, D. Bergmann, S.M. Brady, J. Brkljacic, J. Browse, C. Chapple, V. Colot, S. Cutler, J. Dangel, D. Ehrhardt, J.D. Friesner, W.B. Frommer, E. Grotewold, E. Meyerowitz, J. Nemhauser, M. Nordborg, C. Pikaard, J. Shanklin, C. Somerville, M. Stitt, K.U. Torii, J. Waese, D. Wagner, P. McCourt, 50 years of *Arabidopsis* research: highlights and future directions, *New Phytol.* 209 (2016) 921–944.
- [84] I. Ajjawi, Y. Lu, L.J. Savage, S.M. Bell, R.L. Last, Large-scale reverse genetics in *Arabidopsis*: case studies from the Chloroplast 2010 Project, *Plant Physiol.* 152 (2010) 529–540.
- [85] J. Bergelson, E.S. Buckler, J.R. Ecker, M. Nordborg, D. Weigel, A proposal regarding best practices for validating the identity of genetic stocks and the effects of genetic variants, *Plant Cell* 28 (2016) 606–609.
- [86] R.C. O'Malley, C.C. Barragan, J.R. Ecker, A user's guide to the *Arabidopsis* T-DNA insertion mutant collections, *Methods Mol. Biol.* 1284 (2015) 323–342.
- [87] C.C. Englbrecht, H. Schoof, S. Böhm, Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome, *BMC Genomics* 5 (2004) 39.
- [88] I. Appelhagen, G. Huep, G.H. Lu, G. Strompen, B. Weisshaar, M. Sagasser, Weird fingers: functional analysis of WIP domain proteins, *FEBS Lett.* 584 (2010) 3116–3122.
- [89] B.C. Crawford, G. Ditta, M.F. Yanofsky, The NTT gene is required for transmitting-tract development in carpels of *Arabidopsis thaliana*, *Curr. Biol.* 17 (2007) 1101–1108.
- [90] J.J. Petricka, N.K. Clay, T.M. Nelson, Vein patterning screens and the defectively organized tributaries mutants in *Arabidopsis thaliana*, *Plant J.* 56 (2008) 251–263.
- [91] C. Kühn, W.B. Frommer, A novel zinc finger protein encoded by a couch potato homologue from *Solanum tuberosum* enables a sucrose transport-deficient yeast strain to grow on sucrose, *Mol. Genet.* 247 (1995) 759–763.
- [92] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, *Nat. Protoc.* 10 (2015) 845–858.
- [93] M. van Kregten, S. de Pater, R. Romeijn, R. van Schendel, P.J. Hooykaas, M. Tijsterman, T-DNA integration in plants results from polymerase- θ -mediated DNA repair, *Nat. Plants* 2 (2016) 16164.