Two independent loss-of-function mutations in *anthocyanidin synthase* homeologous genes make sweet basil all green

Itay Gonda[1,*], Mohamad Abu-Abied[1], Chen Adler[1], Renana Milavsky[1,2], Ofir Tal[1], Rachel Davidovich-Rikanati[1], Adi Faigenboim[1], Tali Kahane-Achinoam[1], Alona Shachter[1], David Chaimovitsh[1], Nativ Dudai[1,2,*]

[1] Unit of Aromatic and Medicinal Plants, Newe Ya'ar Research Center, Agricultural Research Organization, Volcani Institute, Ramat-Yishay, Israel

[2] The Hebrew University of Jerusalem, Faculty of Agriculture, Rehovot, Israel

* Corresponding authors: itaygonda@agri.gov.il; nativdud@gmail.com

Abstract

Sweet basil, *Ocimum basilicum* L., is an important culinary herb grown worldwide. Although basil is green, many landraces, breeding lines and exotic cultivars have purple stems and flowers. This anthocyanins pigmentation is unacceptable in the traditional Italian basil. We used the recently published sweet basil genome to map quantitative trait loci (QTL) for flower and stem color in a bi-parental F_2 population. It was found that the pigmentation is governed by a single QTL, harboring an anthocyanidin synthase (ANS) gene. Further analysis revealed that the basil genome harbors two homeologous ANS genes, each carrying a loss-of-function mutation. *ObANS1* carries a 1-bp insertion, and *ObANS2* carries a missense mutation within the active site. In the purple-flower parent, ANS1 is functional and ANS2 carries a nonsense mutation. The functionality of the active allele was validated by complementation in an Arabidopsis ANS mutant. Moreover, we have restored the functionality of the missense-mutated ANS2 using site-directed activation. We found that the non-functional alleles were expressed to similar levels as the functional allele, suggesting polyploids invest futile effort in expressing non-functional genes, harming their superior redundancy. We show here we can harness basil's genomics and genetics to understand the basic mechanism of metabolic traits.

Keywords

A*nthocyanidin synthase*, anthocyanins, genetic mapping, homeologous genes, loss-of-function, sweet basil, tetraploidy

Abbreviations

ANS, anthocyanidin synthase; LG, linkage group; QTL, quantitative trait loci

Introduction

Basil, *Ocimum basilicum* L., of the *Lamiaceae* family, is a leading aromatic crop in agricultural fields and home gardens. It belongs to the genus *Ocimum*, which comprises up to 160 different species (Paton *et al.*, 1999). *O. basilicum* harbors a large diversity among its genotypes depicting unique aromas, leaf sizes and shapes, leaf and stem color, inflorescence color and structure, grew habit and seeds morphology (Dudai and Belanger, 2016). These diverse genotypes have a wide range of uses, primarily as a culinary herb and as a source for essential oils and ornamentals. 'Genovese' basil, the type of basil used in the Italian Pesto sauce, is the most widespread basil grown globally. While commercial 'Genovese' basils are all green, markets also display purple basils or basils with purple stems and flowers.

Anthocyanins are a group of water-soluble pigments conferring purple, red and blue colors in multiple plants (Holton and Cornish, 1995). They exist in almost all plant species except from *Caryophyllales* that accumulate betalains pigments instead (Polturak and Aharoni, 2018; Tanaka *et al.*, 2008). The biosynthesis of anthocyanins is a multistep pathway that starts from the amino acid L-phenylalanine (Fig. 1). Several colorless or yellow intermediates precede the anthocyanins. The first enzyme that generates purple/blue/red pigments is anthocyanidin synthase (ANS; synonym: leucoanthocyanidin dioxygenase, LDOX), which oxygenates the leucoanthocyanidin substrate and generates colored anthocyanidin aglycons (Falcone Ferreyra *et al.*, 2012). Next, multiple sugar substitutions on various residues produce various colorful anthocyanins.

In basil, purple pigmentation is a common feature in the leaves, flowers and stems (Carović-Stanko *et al.*, 2011). While the term purple basil usually refers to the color of the leaves, green-leaf basils with purple flowers and stems are common (Phippen and Simon, 1998). Various anthocyanins were characterized from basil plants, with cultivar being a major factor determining their levels (Flanigan and Niemeyer, 2014; Phippen and Simon, 1998; Prinsi *et al.*, 2019). The accumulation of anthocyanins in basil was also highly dependent on plant age (McCance *et al.*, 2016), peaking at flowering time (Phippen and Simon, 1998). Fall-grown purple basil (cv. Dark Opal) accumulated significantly higher levels of anthocyanins than summer-grown (Nguyen and Niemeyer, 2008). Phippen and Simon (2000) showed that two dominant alleles govern the inheritance of basil color by using a complete diallel cross in segregating F_2 individuals. They documented a high level of instability of purple leaf color seen in spotted or green/purple intermediate phenotypes of the offspring. That instability was

also observed in vegetative cuttings of purple basils, where the position of the cutting influenced its color retaining (Phippen and Simon, 2000).

This project aimed to understand the molecular basis of anthocyanins accumulation in sweet basil. We used a previously-developed F_2 population derived from a cross between a green cultivar and a purple-flower cultivar. Using genotyping-by-sequencing, we showed that a single quantitative trait loci (QTL) governs the color tait. Two mutations in the two ANS homelogous genes were validated. Finally, we raise the question of whether polyploids pay penalties when they express more non-functional genes than diploids, eliminating their hypothesized advantage of redundancy.

Materials and Methods

Plant material

The F_2 mapping population was grown in greenhouse conditions as described in Gonda *et al.* (2022). Flower and stem color were visually evaluated on each F_2 plant in the greenhouse on the emergence of the first flower. For gene expression analysis, 3 plants of 'Perrie' cultivar and 3 plants of 'Cardinal' culivar were grown in open field conditions with drip irrigation. Leaves samples were taken at 10 leaf-pairs stage, from the 7th and 8th pairs. Flowers were sampled after 2 months.

DNA extraction, genotyping by sequencing and association mapping

DNA extraction, GBS libraries construction, SNP calling and association mapping were described in Gonda *et al.* (2022).

Linkage groups determination

Linkage groups (LGs) were built with JoinMap v4.1. Briefly, of the 23,411 SNPs detected, only sites where both parents were homozygous continued the linkage analysis (using Tassel). The data was further filtered with JoinMap, and sites that did not show a disomic distribution of 1:2:1 were filtered out, and only one locus was kept when two loci were > 95% similar. Then, when adjacent sites on the same scaffold were distant less than 2.13 Mbp (on average), only the site with less missing data was kept. Linkage analysis was performed with JoinMap using regression mapping with the Cosambi map function. The parameter used for grouping was recombination frequency from 0.5 to 0.05 with a step of -0.05. LGs were set, and map distances were calculated based on the grouping tree with the regression mapping function. Homeologous LGs were determined considering the BUSCO analysis of Gonda *et al.* (2020) that defined homeologous scaffolds.

RNA-sequencing

RNA sequencing, including RNA extraction, library preparation and gene expression analysis was performed as described in Gonda *et al.* (2020). In addition to the 'Perrie' cultivar, we also used the 'Cardinal' cultivar.

Sequencing *ObANS* genes

Flowers RNA was converted to cDNA using a synthesis kit (PCR Biosystems, https://pcrbio.com/usa/) after DNAse treatment (Thermos Fisher scientific, https://www.thermofisher.com/il/en/home.html). Due to high identity between the homeologous copies, the ANS genes were amplified using a single pair of primers: F 5' ATGGTTGCTTCAATTACGGCA 3'; R 5' CAACTAGATTTATCATCAACCACCACC 3'. The PCR products were cloned into pJET vector (Thermos Fisher scientific) and transformed into DH5α competent cells plated on LB medium containing ampicillin for selection. Ten individual colonies were grown overnight, the plasmids were extracted, and the inserts were sequenced from all colonies to address all possible insert fragments. To restore the functionality of the H292Q mutation of *ObANS1_Perrie* we used the following primers: F 5' CAAAAGCATTCTGCACCGCGCCTCCGTCAA 3'; R 5' TTGACGAGGCCGCGGTGCAGAATGCTTTTG 3'.

Complementation test of *ObANS* genes in Arabidopsis

All *ObANS* genes were inserted into the pBI121 plasmid under the control of the constitutive Cauliflower mosaic virus 35S promoter using the SacI and XbaI restriction enzymes (Thermo-scientific) and the NEBuilder ligation system (NEB). The plasmids were transformed into *Agrobacterium tumefaciens* strain GV3101 using electrotransformation and positive colonies were selected with kanamycin. Agrotransformation of Arabidopsis plants with T-DNA insertion at the *AtANS* gene (SALK_073183, Ohio State University Arabidopsis Biological Resource Center, https://abrc.osu.edu/), were performed using the floral dip technique (Clough and Bent, 1998). The F1 seeds then germinated on a kanamycin-containing MS medium, and seedlings harboring the plasmid were grown, followed by F_2 seeds collection. The F_2 plants were grown on a kanamycin-containing MS medium at 20°C with a 12h photoperiod.

Anthocyanins extraction

Flowers tissues of 3 plants of both 'Perrie' and 'Cardinal' cultivars were collected and flash-frozen in liquid N_2. Samples were then ground to uniform powder. Afterward, 100 mg of

tissue was weighed, 200 µl of 80% methanol was added, and samples were vigorously vortexed. Samples were then sonicated for 20 min at RT followed by 10 min centrifugation at 21,000 *g* at 4°C. The extraction procedure was repeated twice. Finally, the samples were filtered through a syringe filter of 0.22 µm (GHP Membrane, PALL, USA) to new amber vials.

Anthocyanins analysis

Liquid chromatography/time of-flight/mass spectrometry (LC-TOF-MS) analysis was carried out on an Agilent 1290 Infinity series liquid chromatograph coupled with an Agilent 1290 Infinity DAD and Agilent 6530C Accurate Mass quadrupole Time of Flight (qTOF) mass spectrometer (MS) (Agilent Technologies, Santa Clara, USA). Compounds were separated on a Zorbax Extend-C18 Rapid Resolution HT column (2.1 × 50 mm, 1.8 µm; Agilent Technologies). The gradient elution mobile phase consisted of H_2O with 0.1% (v/v) formic acid (eluent A) and acetonitrile containing 0.1% (v/v) formic acid (eluent B). The column was equilibrated with 1% B at a flow rate of 0.3 mL × min^{-1} for 1 min, then increased to 80% B by the following steps: 2-3 min, 20% B; 4-5 min, 30% B; 10-11 min, 50% B; 12-13 min, 80% B. Column was washed with 95 % B at 14 min for 1 min and readjusted to 1% B for 2 minutes. The eluted compounds were subjected to Jet Stream electrospray ionization interface (ESI) operated in positive mode with the following settings: 8 L × min^{-1} gas at 300°C, 35 p.s.i. nebulizer pressure, 10 L × min^{-1} sheath gas at 300°C, capillary voltage (VCap) of 3,000 V, fragmentor to 140 V, and skimmer to 65 V. Data was collected from mass/charge (m/z) ratio of 100-1,700. The flow rate of the mobile phase was 0.3 mL × min^{-1} and the column oven temperature was 30 °C. The main (therefore representative) ions formed in ESI source (mainly $[M]^+$; $[M+H]^+$; $[M+Na]^+$) of target compounds were detected using the 'find compound by formula' function and analyzed by Masshunter qualitative and quantitative analysis software version B.07.00 (Agilent technologies). For untargeted analysis, the platform of MPP (Mass profiler professional, Agilent Technologies) was used. In total, 432 compounds were integrated using molecular feature of which 107 showed > 2 fold change between 'Perrie' and 'Cardinal' samples (Moderate *t*-test, $p < 0.05$). Compounds were annotated using IDBrowser based on exact mass compared to METLIN Metabolite and Chemical Entity Database.

ANS sequence analyses and model predictions

Sequence comparisons and alignments of both nucleotides and amino acids were done with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) using default parameters and

visualized by BoxShade 3.21 (https://embnet.vital-it.ch/software/BOX_form.html). Evolutionary conservation scores were calculated by Consurf (Ashkenazy *et al.*, 2016) for each of the ANS variants. ANS models were predicted using RaptorX (Källberg *et al.*, 2012). AtANS (PDB 1gp6) was used as a template (Wilmouth *et al.*, 2002). For electrostatic surface potential, hydrogens were added using PROPKA (Olsson *et al.*, 2011) followed by calculation of the electrostatic potentials by APBS (Dolinsky *et al.*, 2007). Secondary structure fold prediction was done by HHpred (Söding *et al.*, 2005).

Results

Phenotyping the population for color traits

Many basil landraces and accessions have pink/purple flowers, while most cultivars found on markets have white/green flowers. We used a bi-parental population derived from a cross between the Genovese cultivar 'Perrie' and the ornamental cultivar 'Cardinal' (Dudai *et al.*, 2018) to study the genetic basis for color traits. 'Perrie' has a green stem, green bracts and sepals, and white petals, while 'Cardinal' has a purplish stem, deep purple bracts and sepals and pink petals (Supplementary Fig. S1 at *JXB* online). The purple color results from the presence of four different anthocyanins with various glycoside substitutions, as was observed by LC-TOF analysis (Supplementary Table S1 at *JXB* online). The F_1 plants had purple flowers but not as intense as the 'Cardinal' flower (Supplementary Fig. S1). Next, 173 F_2 segregants were grown and visually scored for the color of the flower parts and the stem. The flower and stem color segregated in the F_2 offspring and displayed varying intensities of purple color. There was an agreement between all non-colored individuals in all traits; offspring with white petals had green stems, bracts and sepals. χ^2 distribution analysis showed that the ratio between the white/green phenotype to the pink/purple phenotype fits a 3:1 single dominant gene inheritance model (Table 1). A look at the intensities of the purple color suggested an intermediate phenotype exists, subdividing the purple phenotype to deep- and light-purple (Table 1; Fig. 2). Although there was no complete agreement on purple intensities among sepals, stem and bracts, contingency analysis within purple segregants only, indicated that the color intensities among these tissues are related to each other (Supplementary Table S2). We have further tested the possibility that the intermediate purple phenotypes resulted from a single gene in an incomplete dominance model. The colors of the bracts, but not the other tissues, were segregated in a 1:2:1 ratio according to the χ^2 probability test (Table 1).

Population's genotyping and linkage groups construction

The population was genotyped with GBS, as described in Gonda *et al.* (2022). Briefly, 23,411 polymorphic sites between 'Perrie' and 'Cardinal' were generated for the mapping population. The heterozygosity levels were 9% for 'Perrie', 33% for 'Cardinal', and the population's mean was 51% ± 0.7.

We next build linkage groups (LGs) and a genetic map based on the SNPs data. To reduce the complexity of the data, we have further filtered the SNPs to have a minimum genotyping depth of 10. The generated ABH genotype (A, 'Perrie'; B, 'Cardinal') contained 11,857 sites (excluding sites where at least one of the parents was heterozygous). The data was further filtered with JoinMap, and sites that did not show a disomic distribution of 1:2:1 were filtered out, and only one locus was kept when two loci were > 95% similar. Then, when adjacent sites on the same scaffold were distant less than 2.13 Mbp (on average), only the site with less missing data was kept. This rigorous filtering resulted in 867 sites spreading over 152 scaffolds. We detected 24 linkage groups, but the average size was only 24 cM (SD = 21.3cM), and the median was 14 cM. Moreover, the SNPs did not order sequentially according to their scaffolds within most of the linkage groups (see examples in Supplementary Fig. S2). Hence, we only divided the scaffolds into LGs and did not continue to pseudomolecules scaffolding. Based on the BUSCO analysis performed by Gonda *et al.* (2020), we have determined homeology between LG couples that were arbitrarily classified into subgenomes A or B (Supplementary Table S3). We also detected chimeric scaffolds that span over 2 or 3 LGs (Supplementary Tables S3-4).

Association mapping of the color traits

Due to the high heterozygosity level of the 'Cardinal' parent and the short genetic sizes of the linkage groups generated, we performed an association analysis rather than QTL mapping. To reduce complexity and false positives, we filtered the data to sites with DP > 15. That resulted in 8,496 polymorphic sites (including sites where one of the parents was heterozygous), which were checked for associations with the various color phenotypes. The results showed that no matter which color trait was tested, three scaffolds were strongly associated with the color traits: 393, 2608 and 7350 (Fig. 3). These scaffolds belong to the same LG (LG 4A), which suggests that the QTL at LG 4A spans over the entire chromosome. Moreover, also scaffold 120, which belongs to LG 4B, showed association with all color traits. We then checked whether the alleles at the QTL (position 7350_2,452,081) also contribute to the incomplete dominance we observed for bracts color phenotype (Fig. 4).

Nonparametric comparisons of each allele-pair using Wilcoxon analysis indicated that the heterozygous status is significantly different from each homozygous status (*p*-value <0.0001). The phenotypic variance explained by this locus is 62%, according to contingency analysis test.

Resolving flower and stem color QTL

The observed intervals of several to dozens Mbp make it challenging to resolve the QTL and find the causative molecular factor using forward genetics only. To overcome that hurdle, we have adopted a reverse genetics approach with candidate genes from the anthocyanins biosynthetic pathway. We used tblastn algorithm to scan the basil genome for the entire anthocyanins biosynthetic pathway genes. We found that the entire pathway is duplicated in the 'Perrie' genome with copies in homeologous scaffolds (Supplementary Table S5). The only gene found in the color QTL scaffolds was the gene encoding for anthocyanidin synthase (ANS) enzyme. One copy of ANS gene was found in scaffold 7350 (termed *ObANS1*) and another copy was found in scaffold 120 (termed *ObANS2*) (Supplementary Table S5). Both genes were predicted to include two exons and are 96% identical in the nucleotide level, the intron however, is more divergent with 85% identities. Two molecular scenarios can explain the no anthocyanins phenotype of 'Perrie' flowers: 1) the expression of both ANS copies is suppressed in 'Perrie' due to *cis*-acting elements, plausibly located in the promoter; 2) 'Perrie' copies of ANS carry mutations in comparison to 'Cardinal' altering their activity. The expression levels of both *ObANS* genes was monitored in the flowers and the leaves of both genotypes using RNA-seq analysis. In both cultivars, both *ANS* genes were highly expressed in the flowers and hardly expressed in the leaves (Fig. 5). *ObANS2* was not differentially expressed between the two cultivars. *ObANS1* showed significantly higher expression levels in 'Cardinal' flowers than in 'Perrie' flowers. Yet, this difference in expression cannot explain the no anthocyanins phenotype of 'Perrie' flowers. Interestingly, both 'Perrie' and 'Cardinal' expressed the entire homeologous sets of anthocyanins biosynthetic genes (Table S6).

***ObANS* enzymes sequence analysis**

Next, we looked for possible causative polymorphic sites between 'Perrie' and 'Cardinal'. For that, we have extracted and sequenced the coding region of the ANS genes from cDNA of 'Perrie' and 'Cardinal' flowers. ANS1_Perrie carries a 1-bp deletion at position 993 bp (Supplementary Fig. S3, red shaded) in comparison to ANS1_Cardinal and ANS2 from both cultivars. ANS1_Cardinal carries a 9-bp deletion after 1,077 bp (Supplementary Fig. S3, light

blue shaded). Several other SNPs were evident among the four ANS genes with 96.3 to 98.1 % identities (Supplementary Table S7). To understand the functional outcome of these deletions and SNPs, we have aligned basil ANS protein sequences along with the protein sequence of the Arabidopsis ANS (AtANS), that its' structure was resolved using X-ray crystallography (Wilmouth *et al.*, 2002). The 9-bp deletion in ObANS1_Cardinal did not seem to cause deletion in functionally important amino acids. All the critical amino acids according to AtANS were conserved in ObANS1_Cardinal (Fig. 6), suggesting it encodes a functional enzyme. In ANS1_Perrie, the 1-bp deletion did not cause a premature stop codon, yet the frameshift causes major variations in the last 58 amino acids of the protein (Fig. 6). That resulted in a protein with only ~83% amino acid identities with other basil ANSs that are 95-98% identical (Supplementary Table S8). According to the crystal structure of AtANS, this arm of the enzyme harbors four amino acid residues that are important to the binding of the substrate. In ANS1_Perrie, these four amino acids were not conserved due to the frameshift plausibly altering its functionality. ObANS2_Cardinal carries a G → T mutation at position 460 (Supplementary Fig. S3, yellow shaded), resulting in a premature stop codon after only 153 amino acids that are 211 amino acids short of the full protein (Fig. 6). Finally, ObANS2_Perrie carries a C → G mutation at position 876 (Supplementary Fig. S3, green shaded), resulting in an amino acid substitution at position 292, replacing histidine with glutamine (Fig. 6). That H292Q mutation is at a histidine residue that is one of the three residues docking the Fe^{++} ion in the enzyme's active site (Wilmouth *et al.*, 2002).

Models for ANS

To evaluate the effect of the mutations in the ANS enzymes of 'Perrie' and to estimate their contribution to the no anthocyanins phenotype, we have used a computational approach to predict their structure and functionality. First, we used Consurf to calculate the structural evolutionary conservation of each sequence residue in the target chain (Supplementary Fig. S4). The lowest score represents the most conserved position in a protein. The His-292 position in ANS proteins has a very low Consurf score of -1.3, indicating a highly evolutionary conserved amino acid. Some positions downstream to residue 311, where the frameshift of ANS1_Perrie occurred, are highly conserved from an evolutionary point of view which is often linked with functionally essential residues.

Next, we have constructed structure models for all four basil ANS enzymes. The Arabidopsis ANS crystal structure (1gp6; resolution of 1.75 Å), which shares ~ 80% amino acid identities with basil ANSs, was used as a template for modeling calculated by RaptorX.

First, we evaluated the influence of the H292Q substitution in ObANS2_Perrie on the active site and its possible effects on binding and activity. The electrostatic potential of the active site's pocket is more negative in ObANS2_Perrie due to the substitution (Supplementary Fig. S5 AB), hence altering substrate and cofactors affinities. As observed in ANS crystallographic structure in the PDB database, the catalytic site populates an Fe^{++} ion as a cofactor, complexed with 2-oxoglutarate (2-OG) and the substrate *t*-dihydroquercetin (DHQ) to form the ANS:Fe(II):2OG:DHQ cluster (The natural substrate of ANS enzyme are leucoanthocyanidins, which are difficult to synthesize, and are unstable in a solution. Hence, the crystal structure of AtANS was resolved with DHQ). The Fe^{++} is coordinated with His-232, His-288, Asp-234, and a bidentate interaction with 2-OG in the 1pg6 structure, which are aligned with His-236, His-292 and Asp-238 respectively in the 'Perrie' and the 'Cardinal' variants (Supplementary Fig. S5 C). Three other coordination axes are held by two water molecules and a single succinate molecule. As indicated in the literature, the residues which mostly form coordination interactions with Fe^{++} ions are: Cys, Asp, Glu, His and Tyr , while Gln is less frequently participates in Fe^{++} coordination. The H292Q substitution might increase the affinity to other ions such as Ca^{++}, K^{++}, Na^{++}, Mg^{++}, Mn^{++} and Co^{++} (Zheng et al., 2008), which in turn would disturb the catalytic functionality and ANS:Fe(II):2OG:DHQ cluster formation.

To evaluate the frameshift effect in ObANS1_Perrie at the C-terminal tail, we used the structural model produced by RaptorX (Supplementary Fig. S6). The model predicts a rigid lobe composed of twisted alpha-helix, partially blocking the catalytic pocket (Supplementary Fig. S6). This kind of rigid structure might contain a hydrophobic core, which stabilizes the enzyme. The electrostatic potential of the frame-shifted C-terminal tail is different compared to the other variants and characterized by a more positive interface. Conservation analysis of the C-terminal raises a few conserved positions (Pro-311, Pro-320, His-341, Lys-345). Phe-334, Ile-338 and Leu-342 participate in DHQ-1 stabilization in the catalytic site and are absent due to frameshift predictably altering the enzyme's catalytic activity.

Complementation test in Arabidopsis

To validate the functionality of sweet basil ANS enzymes and the effect of the different sequences variation, we have introduced them into an Arabidopsis accession with a T-DNA insertion in its' ANS gene. The plants of this Arabidopsis mutant do not accumulate anthocyanins at all in comparison to their Col-0 background (Bowerman *et al.*, 2012). We

have used all four *ObANS* genes and an ObANS2_Perrie with Q → H substitution in position 292 that presumably can fix the causative mutation and restore the enzyme's activity. After floral dip *Agrobacterium*-mediated transformation, F1 plants were selected on Kanamycin-containing MS medium, grown on soil and seeds were harvested. F_2 plants from four independent transformation events were germinated on Kanamycin-containing MS medium and the color of the plants was evaluated. Plants that acquired the *ObANS1_Cardinal* or the *ObANS2_Perrie_Q292H* genes showed an anthocyanin accumulating phenotype (Fig. 7). The anthocyanins were accumulated in the entire plants contrary to the Col-0 WT where they have been accumulated only in the center of the rosette. Contrary, plants acquired *ObANS2_Cardinal*, *ObANS1_Perrie* or *ObANS2_Perrie* showed a green phenotype similar to the phenotype of the Col-0 with T-DNA insertion at the *AtANS* gene.

Discussion

Color trait and anthocyanidin synthase

Anthocyanidin synthase encodes the first purple/blue/red color-committing enzyme in the anthocyanin biosynthetic pathway (Fig. 1). The function and role of ANS were demonstrated in multiple plants, including Arabidopsis (Appelhagen *et al.*, 2011; Pelletier *et al.*, 1997), apple (Szankowski *et al.*, 2009), onion (Kim *et al.*, 2004; Kim *et al.*, 2005) grape (Gollop *et al.*, 2001), and strawberry (Almeida *et al.*, 2007). Multiple plants carry functional mutations in their ANS genes, causing non-colored phenotypes. For example, in pomegranates, it was shown that an insertion in the coding region of *PgANS* gene abolished anthocyanins accumulation in all plant parts (Ben-Simhon *et al.*, 2015). In apple, it was shown that a viable ANS gene is critical for plant survival (Szankowski *et al.*, 2009). Yet, in sweet basil, anthocyanins accumulation did not correlate with drought or salinity stress (Lazarević *et al.*, 2021). Here we showed that two independent mutations in the homeologous genes of basil ANS are responsible for the all-green phenotype of commercial basil. Moreover, a non-functional allele of *ObANS2* was found in a purple-flower basil. The causing mutation is different from the non-functional mutation found in the green basil. Since the active ANS is dominant over the non-active ones, and there is a redundancy resulting from basil tetraploidy, it seems that a strong selection was applied towards green varieties, probably influenced by consumers' demand.

While the mutations in the ANS genes clearly explain the green phenotype of 'Perrie' cultivar, they cannot explain the green leaves phenotype of the 'Cardinal' cultivar. Basils

with purple leaves are commercially available and are very popular as ornamentals (Dudai and Belanger, 2016). They are not limited to a certain chemotype and, although more common in methyl-chavicol accumulating cultivars, purple eugenol and linalool basils exist (Liber *et al.*, 2011; Maggio *et al.*, 2016; Varga *et al.*, 2017). Since 'Cardinal' and all progeny have green leaves, it seems that a different mechanism controls the color trait in the leaves in comparison to the flowers and stem. Phippen and Simon (2000) indicated a two dominant alleles inheritance mechanism for purple basil color. Since no segregant had purple leaves, another gene that alters anthocyanins accumulation in the leaves must exist. This gene might be an active repressor or defective inducer, but it seems that it cause green leaves in both cultivars. The expression of ANS genes is commonly regulated by MYB transcription factors in multiple plants (Xu *et al.*, 2015; Zhang *et al.*, 2014). "Non-purple" plants such as tomato carry a functional ANS gene as was demonstrated in tomato fruits, by exogenous expression of snapdragon transcription factors (Butelli *et al.*, 2008), and by activation-tagged insertion lines (Mathews *et al.*, 2003). We speculate that a similar mechanism is active in 'Cardinal' flowers and stems, and another different mechanism exists in the leaves of purple basil cultivars giving them their unique all-purple phenotype.

Arabidopsis mutants of the ANS gene have brown seeds (Abrahams *et al.*, 2003; Abrahams *et al.*, 2002). However, the seeds of 'Perrie' cultivar that has two non-functional ANS proteins are black as the seeds of 'Cardinal' cultivar and many other sweet basil accessions. That suggests, that in contrary to Arabidopsis, the dark color of basil seeds is not a result of anthocyanins accumulation. Several other phenolics are known to accumulate in seeds and have dark colors, including phlobaphenes accumulated in corn seeds and oxidized proanthocyanidins (Corso *et al.*, 2020; Lepiniec *et al.*, 2006). Yet, it seems that multiple metabolites contribute to the final color of the testa, as has been documented in *Brassica* (Yu, 2013). A detailed liquid-chromatography – mass-spectrometry analysis of basil seed coat may answer the question of which pigments contribute to the dark appearance of the testa in the lack of anthocyanins.

Basil genomics and genetics

All color traits examined in this study showed disomic segregation with a dominant or incomplete dominance single gene inheritance model similar to the fusarium resistance trait (Gonda *et al.*, 2022). That was also supported by the molecular findings of only a single functional ANS gene in 'Cardinal' cultivar. Moreover, the most proximate SNP to *ObANS1*, 7350_2,452,081, showed a disomic 1:2:1 distribution. On the contrary, the association

mapping indicated a QTL on scaffold 120, where *ObANS2* is located. However, both *ObANS2_Perrie* and *ObANS2_Caridnal* are non-functional. Hence, the QTL observed at scaffold 120 can probably be explained by erroneous mapping of homeoloug reads to this scaffold instead of scaffold 7350. Support for this hypothesis can be found in the high level of heterozygosity detected in the 'Cardinal' parent, which was discussed by Gonda *et al.* (2022).

The large size of the QTL across the entire LG4A is in accordance with the small genetic distances found. It is currently unclear whether the large QTL is an artifact caused by genotyping errors, scaffolding errors, basil tetraploidy that causes a bioinformatic hurdle, or a real biology phenomenon. The parental lines might be too distant from each other, causing low recombination frequencies, resulting in small genetic distances and large QTLs. Large QTLs were observed in other corps, especially when located next to the centromere (Galpaz *et al.*, 2018; Gao *et al.*, 2019). Considering the very narrow QTL found for Fusarium wilt resistance in the same mapping population (Gonda *et al*., 2022), the later explanation should be considered.

Expression of genes coding non-functional proteins

Another aspect of the current research is the high expression of genes encoding non-functional proteins. Both 'Perrie' and 'Cardinal' express inactive versions of the ANS gene. About 65% of the expressed *ObANS* RNA encoded mutant alleles. We have also observed an expression of mutated genes in the phenylpropenes biosynthesis pathway (Gonda et al., 2020). Phenotypically, polyploids can "allow" themselves to express mutated genes more than diploids since they have gene redundancy, as in the case of the flower color of 'Cardinal'. Moreover, the selection against these mutated versions would be smaller because of this redundancy. However, this unnecessary expression and probable translation is futile and has energy costs. When dealing with biosynthetic pathways, such as anthocyanins biosynthesis, this cost is multiplied by the number of the genes involved. The fact that 'Perrie' was expressing the entire biosynthetic pathway in most of the cases of both homelogous caused even a greater futile expression effort. There is a long debate on the beneficiary of being a polyploid and its role in evolution and agriculture (Chen, 2010; Madlung, 2013). It has been suggested that polyploidy does not contribute to increased speciation and is an evolutionary dead-end (Mayrose *et al.*, 2011; Wood *et al.*, 2009). Yet, within domesticate plants, Salman-Minkov *et al.* (2016) showed that polyploidy promoted diversity and allowed wild polyploids to cope with agricultural niches and conditions.

Whatever will be the final answer in this debate, the degree of expression of genes encoding non-functional proteins and their possible energy cost should be considered. Whether it is beneficial for an organism to keep biosynthesis pathways "alive" at the expense of energy cost is another open question to be answered.

Supplementary data

Table S1. Anthocyanins composition in 'Cardinal' cultivar vs. 'Perrie' cultivar.

Table S2. Contingency analysis of the phenotype within the purple segregants only.

Table S3. Scaffolds distribution to linkage groups.

Table S4. Chimeric scaffolds.

Table S5. Distribution of anthocyanins biosynthesis genes across sweet basil subgenomes.

Table S6. Expression of anthocyanins biosynthesis genes in 'Perrie' and 'Cardinal' flowers and leaves.

Table S7. Percentage identities of nucleotides of *ObANSs*.

Table S8. Percentage identities of amino acids of ObANS proteins and AtANS.

Fig. S1 Schematic illustration of the research population.

Fig. S2. Map charts of LG 1 and LG 4.

Fig. S3. Multiple sequence alignment of sweet basil ANS genes.

Fig. S4. Conserved residues analysis of ANS protein.

Fig. S5. A 3-D model structure of the active site of ObANS2_Perrie.

Fig. S6. A 3-D model structure for ObANS1_Perrie.

Acknowledgments

The authors wish to thank lab and team members of the Unit of Aromatic and Medicinal Plants in Newe-Ya'ar Research Center in the past and present. Special thanks to Prof. Asaph Aharoni and Dr. Elad Oren for helpful discussion and brainstorming.

Author contribution

RM, CA, AS, and MAA performed field and molecular biology experiments. DC and TKA performed and supervised field experiments. RDK and CA performed biochemistry analysis.

AF, OT and IG performed the bioinformatics analyses as well as the data curation. IG wrote the manuscript. IG and ND conceptualize the research and review the manuscript.

Conflict of interests

Authors declare no conflict of interests.

Data availability

The raw data of the GBS is available through NCBI short reads archive (SRA) under project number: PRJNA836178.

References

Abrahams S, Lee E, Walker AR, Tanner GJ, Larkin PJ, Ashton AR. 2003. The Arabidopsis TDS4 gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. The Plant Journal **35**, 624-636.

Abrahams S, Tanner GJ, Larkin PJ, Ashton AR. 2002. Identification and biochemical characterization of mutants in the proanthocyanidin pathway in Arabidopsis. Plant Physiology **130**, 561-576.

Almeida JRM, D'Amico E, Preuss A, Carbone F, de Vos CHR, Deiml B, Mourgues F, Perrotta G, Fischer TC, Bovy AG, Martens S, Rosati C. 2007. Characterization of major enzymes and genes involved in flavonoid and proanthocyanidin biosynthesis during fruit development in strawberry (*Fragaria × ananassa*). Archives of Biochemistry and Biophysics **465**, 61-71.

Appelhagen I, Jahns O, Bartelniewoehner L, Sagasser M, Weisshaar B, Stracke R. 2011. Leucoanthocyanidin dioxygenase in *Arabidopsis thaliana*: characterization of mutant alleles and regulation by MYB–BHLH–TTG1 transcription factor complexes. Gene **484**, 61-68.

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Research **44**, W344-W350.

Ben-Simhon Z, Judeinstein S, Trainin T, Harel-Beja R, Bar-Ya'akov I, Borochov-Neori H, Holland D. 2015. A "white" anthocyanin-less pomegranate (*Punica granatum* L.) caused by an insertion in the coding region of the leucoanthocyanidin dioxygenase (LDOX; ANS) gene. PloS One **10**, e0142777.

Bowerman PA, Ramirez MV, Price MB, Helm RF, Winkel BSJ. 2012. Analysis of T-DNA alleles of flavonoid biosynthesis genes in Arabidopsis ecotype Columbia. BMC Research Notes **5**, 485.

Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Schijlen EG, Hall RD, Bovy AG, Luo J, Martin C. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature Biotechnology **26**, 1301-1308.

Carović-Stanko K, Šalinović A, Grdiša M, Liber Z, Kolak I, Satovic Z. 2011. Efficiency of morphological trait descriptors in discrimination of Ocimum basilicum L. accessions. Plant Biosystems **145**, 298-305.

Chen ZJ. 2010. Molecular mechanisms of polyploidy and hybrid vigor. Trends in Plant Science **15**, 57-71.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium* -mediated transformation of *Arabidopsis thaliana*. Plant Journal **16**, 735-743.

Corso M, Perreau F, Mouille G, Lepiniec L. 2020. Specialized phenolic compounds in seeds: structures, functions, and regulations. Plant Science **296**, 110471.

Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Research **35**, W522-W525.

Dudai N, Belanger FC. 2016. Aroma as a factor in the breeding process of fresh herbs – the case of basil. In: Havkin-Frenkel D, Dudai N, eds. *Biotechnology in Flavor Production*, 32-61.

Dudai N, Li G, Shachter A, Belanger F, Chaimovitsh D. 2018. Heredity of phenylpropenes in sweet basil (*Ocimum basilicum* L.) chemotypes and their distribution within an F_2 population. Plant Breeding **137**, 443-449.

Falcone Ferreyra ML, Rius S, Casati P. 2012. Flavonoids: biosynthesis, biological functions, and biotechnological applications. Frontiers in Plant Science **3**, 222.

Flanigan PM, Niemeyer ED. 2014. Effect of cultivar on phenolic levels, anthocyanin composition, and antioxidant properties in purple basil (*Ocimum basilicum* L.). Food Chemistry **164**, 518-526.

Galpaz N, Gonda I, Shem-Tov D, Barad O, Tzuri G, Lev S, Fei Z, Xu Y, Mao L, Jiao C, Harel-Beja R, Doron-Faigenboim A, Tzfadia O, Bar E, Meir A, Sa'ar U, Fait A, Halperin E, Kenigswald M, Fallik E, Lombardi N, Kol G, Ronen G, Burger Y, Gur A, Tadmor Y, Portnoy V, Schaffer AA, Lewinsohn E, Giovannoni JJ, Katzir N. 2018. Deciphering genetic factors that determine melon

fruit-quality traits using RNA-Seq-based high-resolution QTL and eQTL mapping. Plant Journal **94**, 169-191.

Gao L, Gonda I, Sun H, Ma Q, Bao K, Tieman DM, Burzynski-Chang EA, Fish TL, Stromberg KA, Sacks GL, Thannhauser TW, Foolad MR, Diez MJ, Blanca J, Canizares J, Xu Y, van der Knaap E, Huang S, Klee HJ, Giovannoni JJ, Fei Z. 2019. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nature Genetics **51**, 1044-1051.

Gollop R, Farhi S, Perl A. 2001. Regulation of the leucoanthocyanidin dioxygenase gene expression in *Vitis vinifera*. Plant Science **161**, 579-588.

Gonda I, Faigenboim A, Adler C, Milavski R, Karp M-J, Shachter A, Ronen G, Baruch K, Chaimovitsh D, Dudai N. 2020. The genome sequence of tetraploid sweet basil, *Ocimum basilicum* L., provides tools for advanced genome editing and molecular breeding. DNA Research **27**, dsaa027.

Gonda I, Milavski R, Adler C, Abu-Abied M, Tal O, Faigenboim A, Chaimovitsh D, Dudai N. 2022. Genome-based high-resolution mapping of fusarium wilt resistance in sweet basil. Plant Science, 111316.

Holton TA, Cornish EC. 1995. Genetics and Biochemistry of Anthocyanin Biosynthesis. The Plant Cell **7**, 1071-1083.

Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. 2012. Template-based protein structure modeling using the RaptorX web server. Nature Protocols **7**, 1511-1522.

Kim S, Binzel ML, Yoo KS, Park S, Pike LM. 2004. Pink (P), a new locus responsible for a pink trait in onions (*Allium cepa*) resulting from natural mutations of anthocyanidin synthase. Molecular Genetics and Genomics **272**, 18-27.

Kim S, Jones R, Yoo K-S, Pike LM. 2005. The *L* locus, one of complementary genes required for anthocyanin production in onions (*Allium cepa*), encodes anthocyanidin synthase. Theoretical and Applied Genetics **111**, 120-127.

Lazarević B, Šatović Z, Nimac A, Vidak M, Gunjača J, Politeo O, Carović-Stanko K. 2021. Application of phenotyping methods in detection of drought and salinity stress in basil (*Ocimum basilicum* L.). Frontiers in Plant Science **12**, 629441.

Lepiniec L, Debeaujon I, Routaboul J-M, Baudry A, Pourcel L, Nesi N, Caboche M. 2006. Genetics and biochemistry of seed flavonoids. Annual Review of Plant Biology **57**, 405-430.

Liber Z, Carović-Stanko K, Politeo O, Strikić F, Kolak I, Milos M, Satovic Z. 2011. Chemical characterization and genetic relationships among *Ocimum basilicum* L. cultivars. Chemistry & Biodiversity **8**, 1978-1989.

Madlung A. 2013. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity **110**, 99-104.

Maggio A, Roscigno G, Bruno M, De Falco E, Senatore F. 2016. Essential-oil variability in a collection of *Ocimum basilicum* L. (basil) cultivars. Chemistry & Biodiversity **13**, 1357-1368.

Mathews H, Clendennen SK, Caldwell CG, Liu XL, Connors K, Matheis N, Schuster DK, Menasco DJ, Wagoner W, Lightner J, Wagner DR. 2003. Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. The Plant Cell **15**, 1689-1703.

Mayrose I, Zhan Shing H, Rothfels Carl J, Magnuson-Ford K, Barker Michael S, Rieseberg Loren H, Otto Sarah P. 2011. Recently formed polyploid plants diversify at lower rates. Science **333**, 1257-1257.

McCance KR, Flanigan PM, Quick MM, Niemeyer ED. 2016. Influence of plant maturity on anthocyanin concentrations, phenolic composition, and antioxidant properties of 3 purple basil (*Ocimum basilicum* L.) cultivars. Journal of Food Composition and Analysis **53**, 30-39.

Nguyen PM, Niemeyer ED. 2008. Effects of nitrogen fertilization on the phenolic composition and antioxidant properties of basil (*Ocimum basilicum* L.). Journal of Agricultural and Food Chemistry **56**, 8685-8691.

Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. 2011. PROPKA3: Consistent treatment of internal and surface residues in empirical p*K*a predictions. Journal of Chemical Theory and Computation **7**, 525-537.

Paton A, Harley RM, Harley MM. 1999. *Ocimum*: an overview of classification and relationships. In: Hiltunen R, Holm Y, eds. *Basil*. The Netherlands: Harwood Academic, 1-32.

Pelletier MK, Murrell JR, Shirley BW. 1997. Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in Arabidopsis (further evidence for differential regulation of "early" and "late" genes). Plant Physiology **113**, 1437-1445.

Phippen WB, Simon JE. 1998. Anthocyanins in basil (*Ocimum basilicum* L.). Journal of Agricultural and Food Chemistry **46**, 1734-1738.

Phippen WB, Simon JE. 2000. Anthocyanin inheritance and instability in purple basil (*Ocimum basilicum* L.). Journal of Heredity **91**, 289-296.

Polturak G, Aharoni A. 2018. "La Vie en Rose": Biosynthesis, sources, and applications of betalain pigments. Molecular Plant **11**, 7-22.

Prinsi B, Morgutti S, Negrini N, Faoro F, Espen L. 2019. Insight into composition of bioactive phenolic compounds in leaves and flowers of green and purple basil. Plants **9**, 22.

Salman-Minkov A, Sabath N, Mayrose I. 2016. Whole-genome duplication as a key factor in crop domestication. Nature Plants **2**, 16115.

Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research **33**, W244-W248.

Szankowski I, Flachowsky H, Li H, Halbwirth H, Treutter D, Regos I, Hanke M-V, Stich K, Fischer TC. 2009. Shift in polyphenol profile and sublethal phenotype caused by silencing of anthocyanidin synthase in apple (*Malus* sp.). Planta **229**, 681-692.

Tanaka Y, Sasaki N, Ohmiya A. 2008. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. The Plant Journal **54**, 733-749.

Varga F, Carović-Stanko K, Ristić M, Grdiša M, Liber Z, Šatović Z. 2017. Morphological and biochemical intraspecific characterization of *Ocimum basilicum* L. Industrial Crops and Products **109**, 611-618.

Wilmouth RC, Turnbull JJ, Welford RWD, Clifton IJ, Prescott AG, Schofield CJ. 2002. Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*. Structure **10**, 93-103.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences **106**, 13875-13879.

Xu W, Dubos C, Lepiniec L. 2015. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends in Plant Science **20**, 176-185.

Yu C-Y. 2013. Molecular mechanism of manipulating seed coat coloration in oilseed Brassica species. Journal of Applied Genetics **54**, 135-145.

Zhang Y, Butelli E, Martin C. 2014. Engineering anthocyanin biosynthesis in plants. Current Opinion in Plant Biology **19**, 81-90.

Tables

Table 1. Observed and expected colors phenotype in the research population

Tissue	Observed Phenotypes	Observations	Expected Observations	Tested Ratio	Genetic Model	χ^2 Square	P - value	DF
Stem / Sepals / Bracts	purple:green	136:37	130:43	3:1	single dominant	1.24	0.26	1
Sepals	pink:white	136:37	130:43	3:1	single dominant	1.24	0.26	1
Stem	deep purple:light purple:green	92:44:37	43:87:43	1:2:1	incomplete dominance	76.73	<0.0001	2
Sepelas	deep purple:light purple:green	69:67:37	43:87:43	1:2:1	incomplete dominance	17.48	<0.0001	2
Bracts	deep purple:light purple:green	48:88:37	43:87:43	1:2:1	incomplete dominance	1.45	0.48	2

Figures legends

Fig. 1 Biosynthesis pathway of anthocyanins. PAL, phenylalanine ammonia-lyase; 4CL, *p*-coumarate CoA ligase; C4H, coumarate 4-hydroxylase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; UFGT, UDP-glucose flavonoid 3-*O*-glucosyltransferase. Enzyme names are in italic letters.

Fig. 2 Color phenotype within the F_2 population. Three representative color phenotypes of the F_2 offspring: green with white sepals (left), light purple with pink petals (middle), and deep-purple with purple petals (right).

Fig. 3 Association mapping of basil color traits across sweet basil genome scaffolds. Manhattan plots of the QTL for the color of: A. Bract leaves, B. Sepals, C. Stem, D. Petals. The associations of the traits and SNP within the F_2 population were calculated using Tassel with GLM algorithm. Different colors and symbols represent the different scaffolds. Numbers represent: 1, scaffold 120; 2, scaffold 2608; scaffold 393; 4, scaffold 7350.

Fig. 4 Contingency analysis of the bract leaves color phenotype vs. the allelic composition at site 7350_2452081. The width of each tile is proportional to the number of segregants with a given allelic composition in scaffold 7350 at position 2,452,081 bp. The height of each tile represents the fraction of the segregants with a given color with a given allelic composition. n=173.

Fig. 5 Expression levels of ANS genes in sweet basil. The expression of *ObANS1* and *ObANS2* in the parental lines, 'Perrie' and 'Cardinal', in the flowers and leaves was determined using RNA-seq. Values are the mean of 3 biological repeats ± SEM.

Fig. 6 Multiple sequence alignment of anthocyanidin synthase proteins from sweet basil and Arabidopsis. Comparison between the amino acid sequences of basil ANS enzymes from 'Perrie' and 'Cardinal'. Residues conserved in more than four sequences are black shaded, and similar residues are gray shaded. Shapes and colors represent conserved residues based on the crystallographic structure of AtANS (Wilmouth et al., 2002). AtANS, Arabidopsis thaliana, accession number NP_194019.1; DHQ, dihydroquercetin.

Fig. 7 Complementation of the no anthocyanin phenotype in Arabidopsis by *ObANS* genes. The coding sequences of the four *ObANS* were extracted from cDNA, and cloned under the 35S promoter. The Q292H substitution in ObANS2_Perrie was performed using specific primers. The genes were transformed into Arabidopsis plants (Col-0 background) with a T-DNA insertion at the *AtANS* gene. The pictures depict F_2 transgenes grown in MS medium with kanamycin at 20°C with continuous light. Each transgenic plant is representative of an independent transformation event.

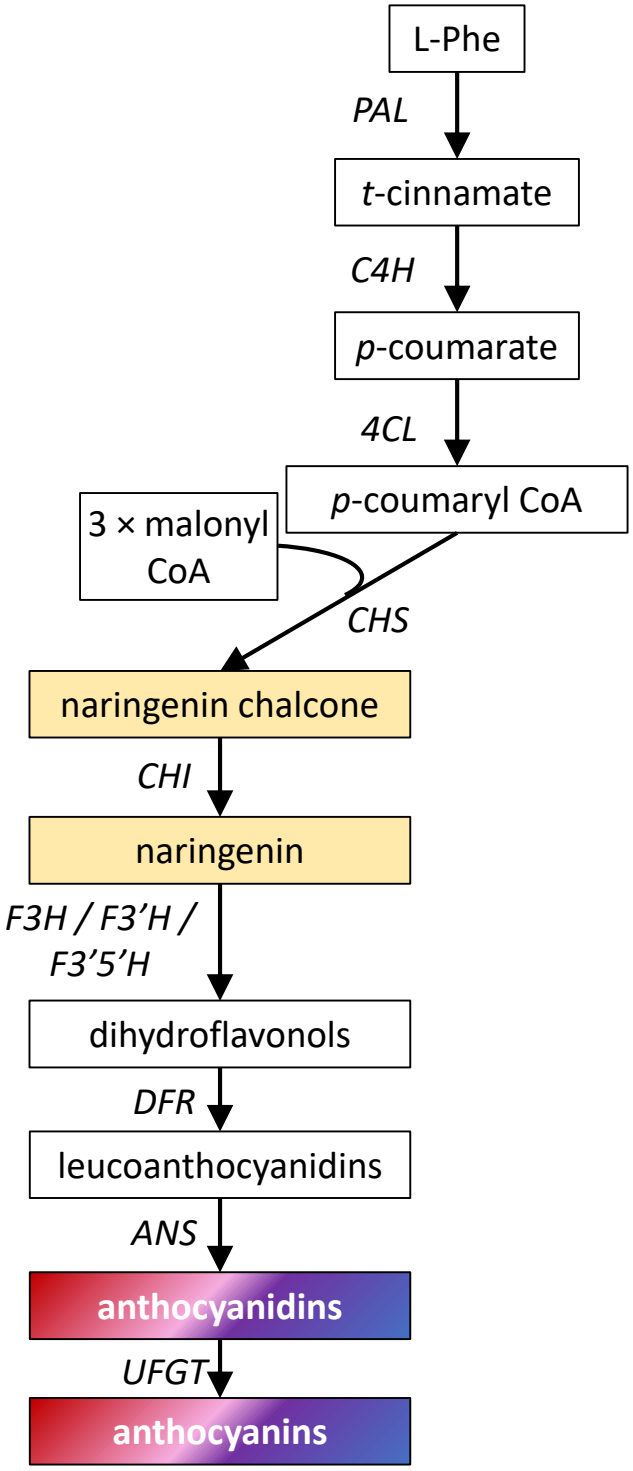

Fig. 1 Biosynthesis pathway of anthocyanins. PAL, phenylalanine ammonia lyase; 4CL, p-coumarate CoA ligase; C4H, coumarate 4-hydroxylase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; UFGT, UDP-glucose flavonoid 3-O-glucosyltransferase. Enzyme names are in italic letters.

Figure 2. Color phenotype within the F2 population.

Three representative color phenotypes of the F2 offspring: green with white sepals (left), light purple with pink petals (middle), deep purple with purple petals (right).

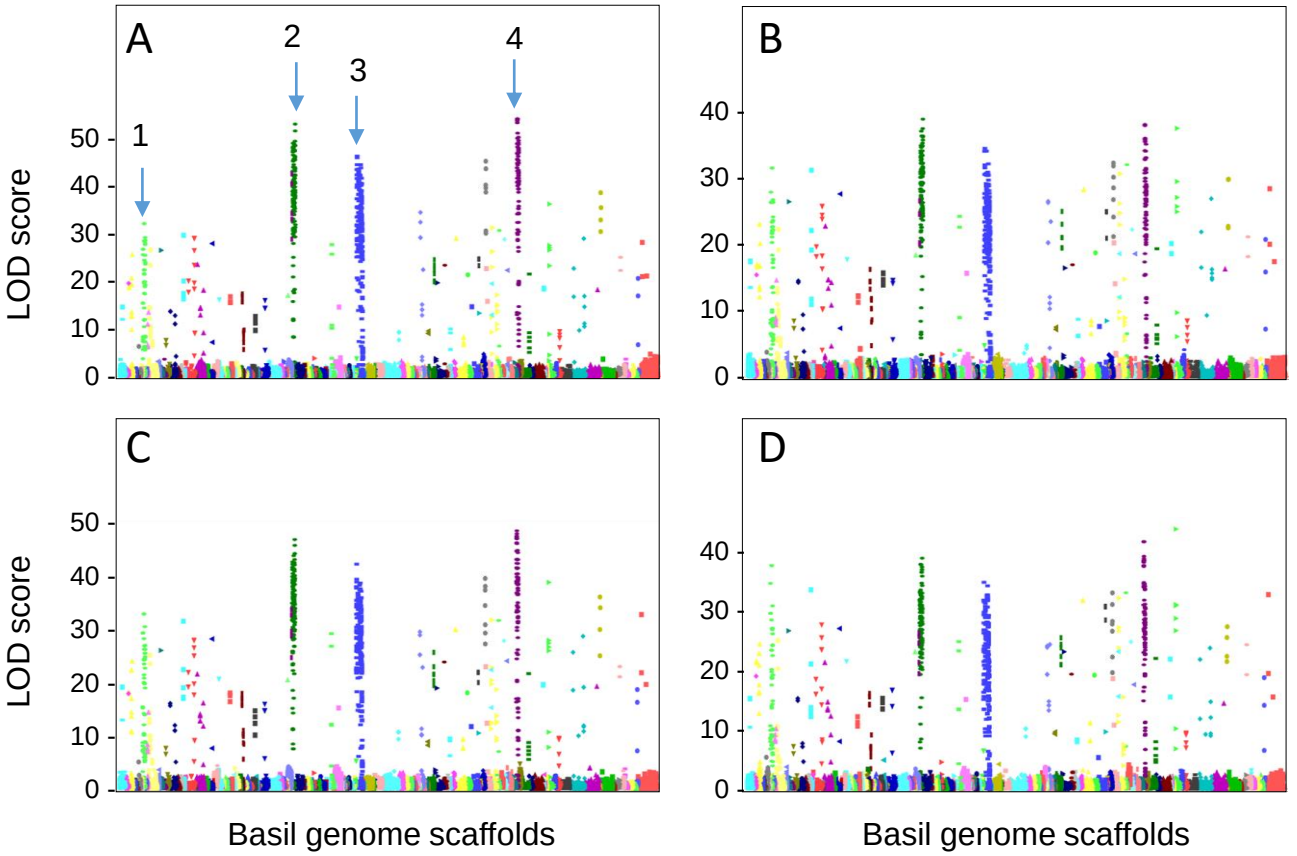

Fig. 3. Association mapping of basil color traits across sweet basil genome scaffolds. Manhattan plots of the QTL for the color of: A. Bract leaves, B. Sepals, C. Stem, D. Petals. The associations of the traits and SNP within the F2 population were calculated using Tassel with GLM algorithm. Different colors and symbols represent the different scaffolds. Numbers represent: 1, scaffold 120; 2, scaffold 2608; scaffold 393; 4, scaffold 7350.

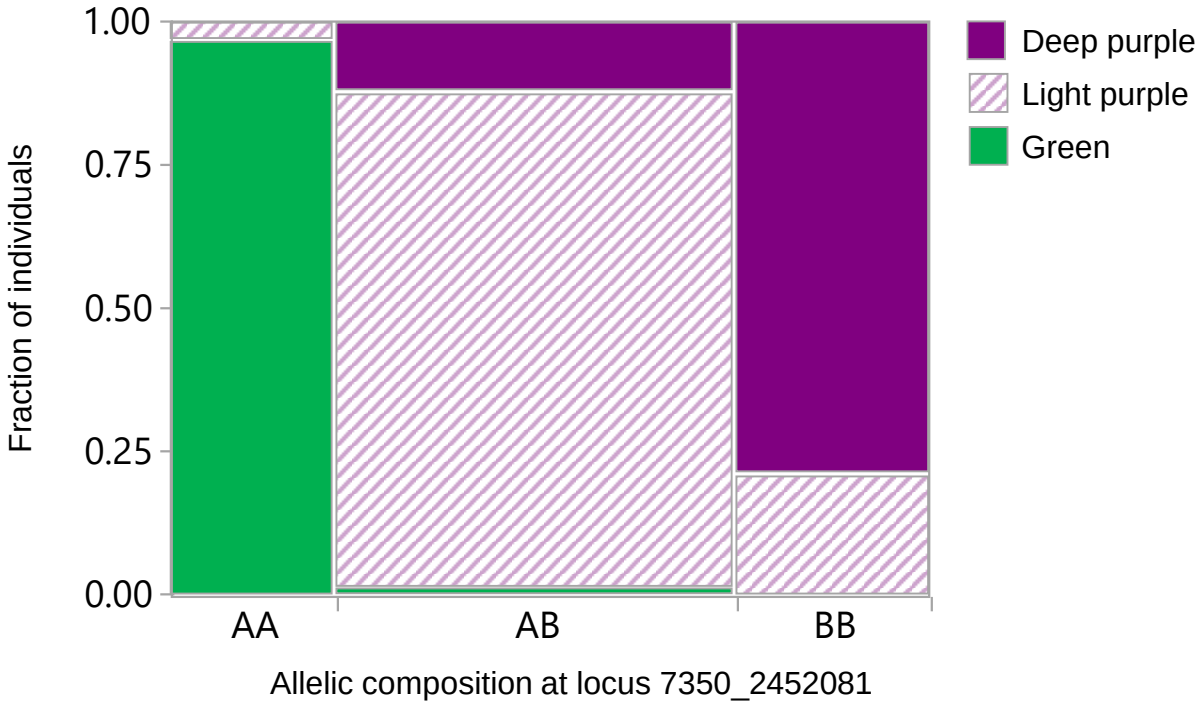

Figure 4. Contingency analysis of the bract leaves color phenotype vs. the allelic composition at site 7350_2452081. The width of each tile is proportional to the number of segregants with a given allelic composition in scaffold 7350 at position 2,452,081 bp. The height of each tile represents the fraction of the segregants with a given color with a given allelic composition. n=173.

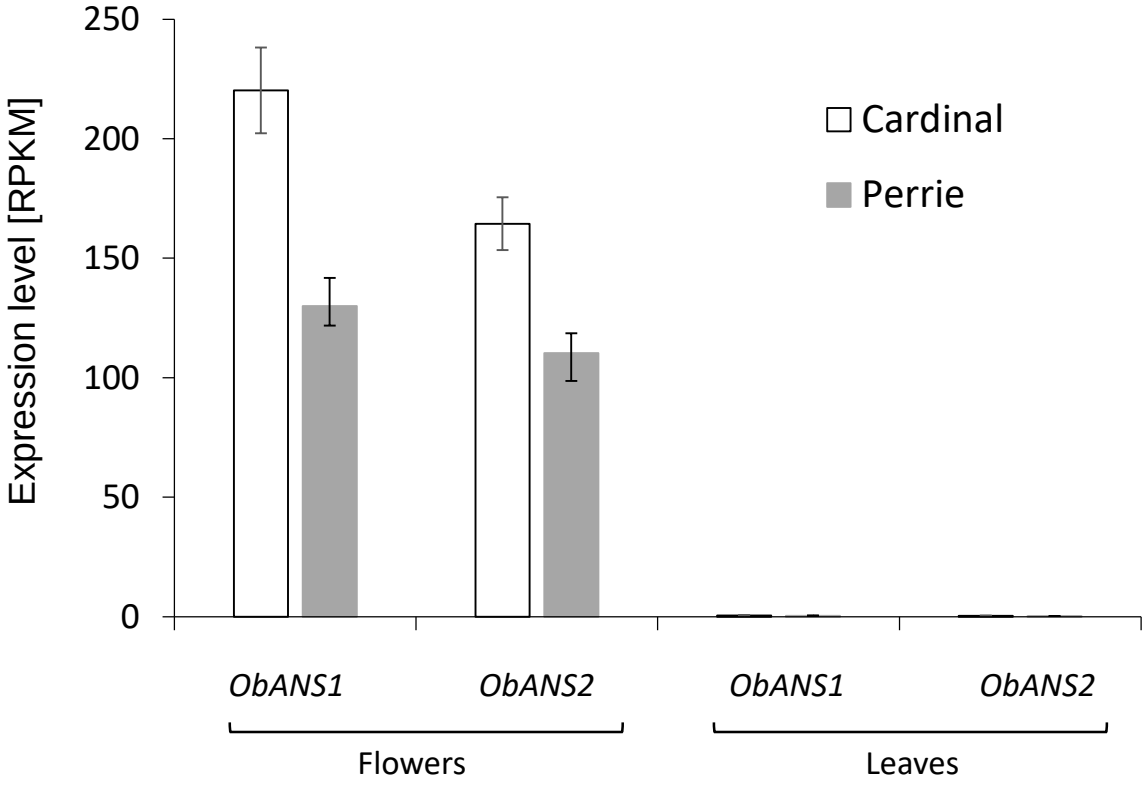

Figure 5. Expression levels of ANS genes in sweet basil. The expression of *ObANS1* and *ObANS2* in the parental lines, 'Perrie' and 'Cardinal', in the flowers and leaves was determined using RNA-seq. Values are mean of 3 biological repeats ± SEM.

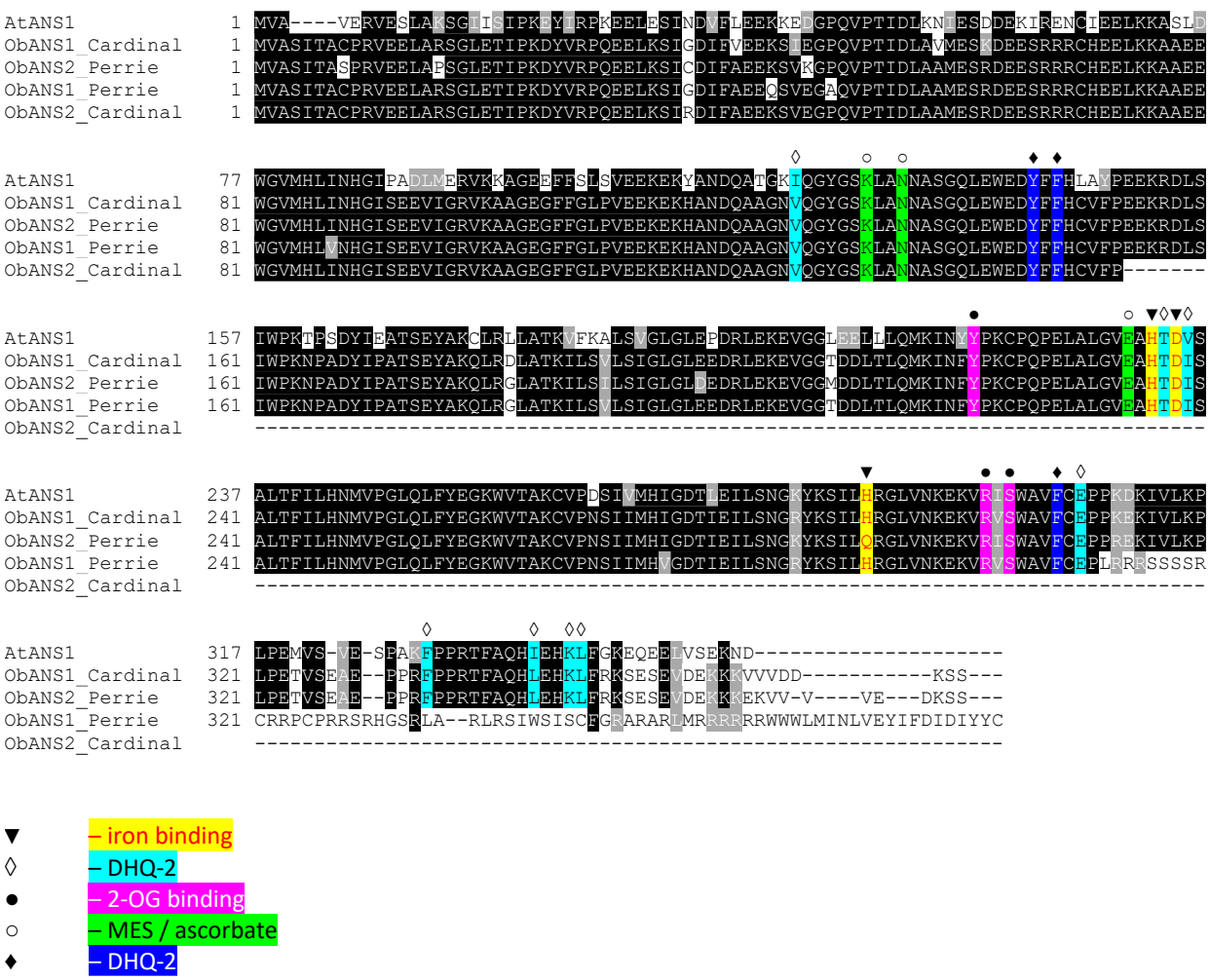

Figure 6. Multiple sequence alignment of anthocyanidin synthase proteins from sweet basil and Arabidopsis. Comparison between the amino acid sequences of basil ANS enzymes from 'Perrie' and 'Cardinal'. Residues conserved in more than four sequence are black shaded and similar residues are gray shaded. Shapes and colors represent conserved residues based on the crystallographic structure of AtANS (Wilmouth et al., 2002). AtANS, *Arabidopsis thaliana*, accession number NP_194019.1.

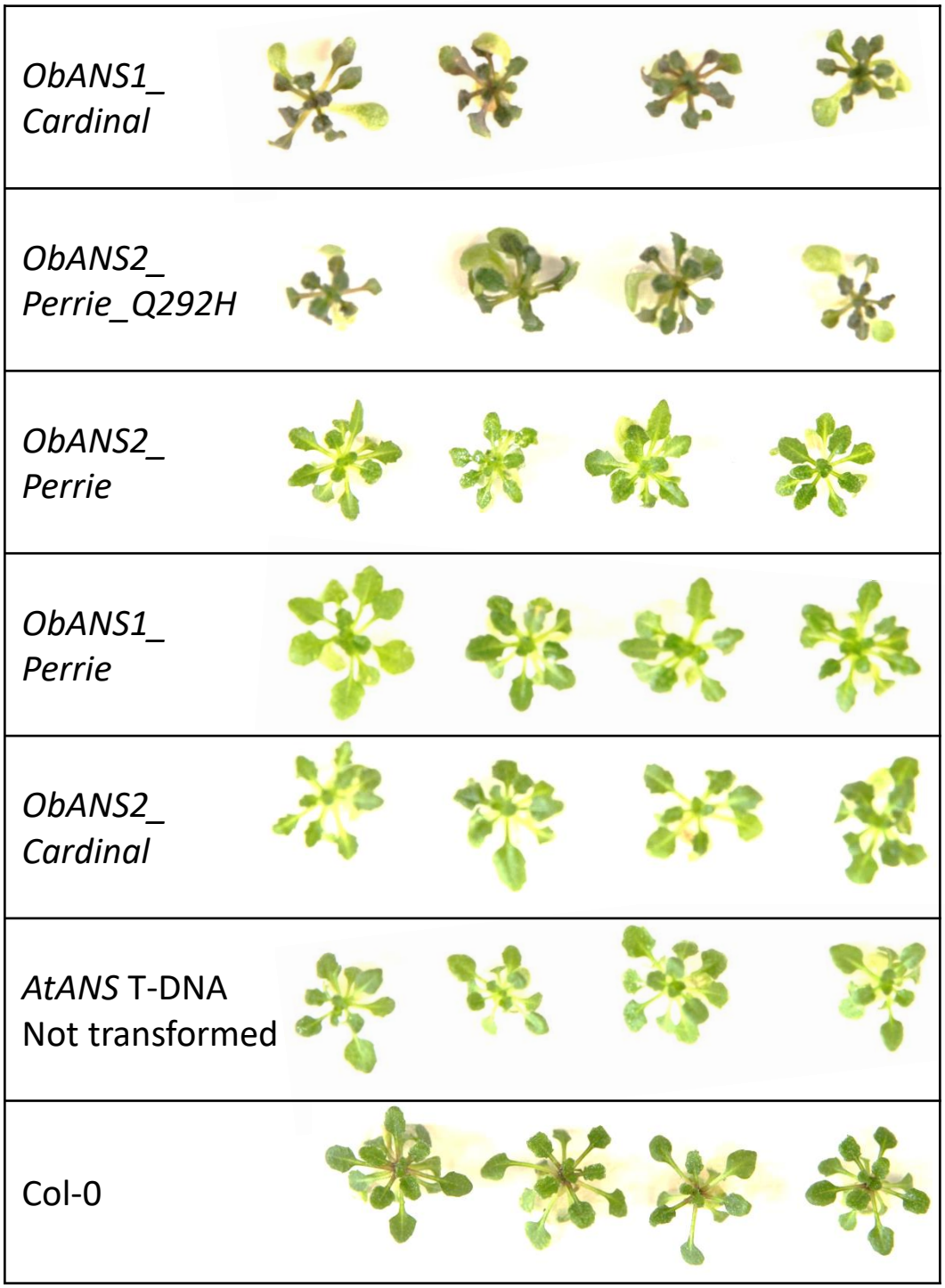

Figure 7. **Complementation of the no anthocyanin phenotype in Arabidopsis by *ObANS* genes.** The coding sequences of the four *ObANS* were extracted from cDNA, and cloned under the 35S promoter. The Q292H substitution in ObANS2_Perrie was performed using specific primers. The genes were transformed into Arabidopsis plants (Col-0 background) with a T-DNA insertion at the *AtANS* gene. The pictures depict F2 transgenes grown in MS medium with kanamycin at 20°C with continuous light. Each transgenic plant is representative of an independent transformation event.