



# Evolutionary and functional analyses of the 2-oxoglutarate-dependent dioxygenase genes involved in the flavonoid biosynthesis pathway in tobacco

Zhong Wang<sup>1</sup> · Shanshan Wang<sup>1</sup> · Mingzhu Wu<sup>1</sup> · Zefeng Li<sup>1</sup> · Pingping Liu<sup>1</sup> · Feng Li<sup>1</sup> · Qiansi Chen<sup>1</sup> · Aiguo Yang<sup>2</sup> · Jun Yang<sup>1</sup>

Received: 13 June 2018 / Accepted: 23 September 2018 / Published online: 6 October 2018  
© The Author(s) 2018

## Abstract

**Main conclusion** This study illustrates the differences in the gene structure of 2-oxoglutarate-dependent oxygenase involved in flavonoid biosynthesis (2ODD-IFB), and their potential roles in regulating tobacco flavonoid biosynthesis and plant growth.

Flavonol synthase (FLS), anthocyanidin synthase (ANS), and flavanone 3 $\beta$ -hydroxylase belong to the 2-oxoglutarate-dependent (2ODD) oxygenase family, and each performs crucial functions in the biosynthesis of flavonoids. We identified two *NtFLS* genes, two *NtANS* genes, and four *NtF3H* genes from *Nicotiana tabacum* genome, as well as their homologous genes in the *N. sylvestris* and *N. tomentosiformis* genomes. Our phylogenetic analysis indicated that these three types of genes split from each other before the divergence of gymnosperms and angiosperms. *FLS* evolved faster in the eudicot plants, whereas *ANS* evolved faster in the monocot plants. Gene structure analysis revealed two fragment insertions occurred at different times in the intron one position of tobacco *FLS* genes. Homologous protein modeling revealed distinct structures in the N terminus of the tobacco 2ODD oxygenases. We found that the expression patterns of genes encoding tobacco 2ODD oxygenases in flavonoids biosynthesis (2ODD-IFB) did not determine the accumulation patterns of flavonoids among various tobacco tissues, but strongly affected the concentration of flavonoids in the tissues, where they were biosynthesized. More carbon resource flowed to the flavonol biosynthesis when *NtANS* gene was silenced, otherwise more anthocyanidin accumulated when *NtFLS* gene was repressed. This study illustrates the 2ODD-IFB gene structure evolution, differences among their protein structures, and provides a foundation for regulating plant development and altering flavonoid content and/or composition through the manipulation of plant 2ODD-IFB genes.

**Keywords** Abiotic stress · Flavonoids · Intron length · Phylogeny · Tissue-specific expression

## Abbreviations

ANS	Anthocyanidin synthase	FLS	Flavonol synthase
DPBA	Diphenylborinic anhydride	2ODD	2-Oxoglutarate dependent
F3H	Flavanone 3 $\beta$ -hydroxylase	2ODD-IFB	2ODD involved in flavonoid biosynthesis
		VIGS	Virus-induced gene silencing

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00425-018-3019-2>) contains supplementary material, which is available to authorized users.

✉ Jun Yang  
yangjun@ztri.com.cn

<sup>1</sup> China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China

<sup>2</sup> Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao 266101, China

## Introduction

Flavonoids are polyphenolic secondary metabolites with C6–C3–C6 carbon backbones that are biosynthesized via the phenylpropanoid pathway (Winkel-Shirley 2001). There are approximately 10,000 known flavonoids in plants; these are classified into different subgroups depending on the organization and modifications of the three-ring structure (Harborne and Williams 2000; Williams and Grayer 2004).

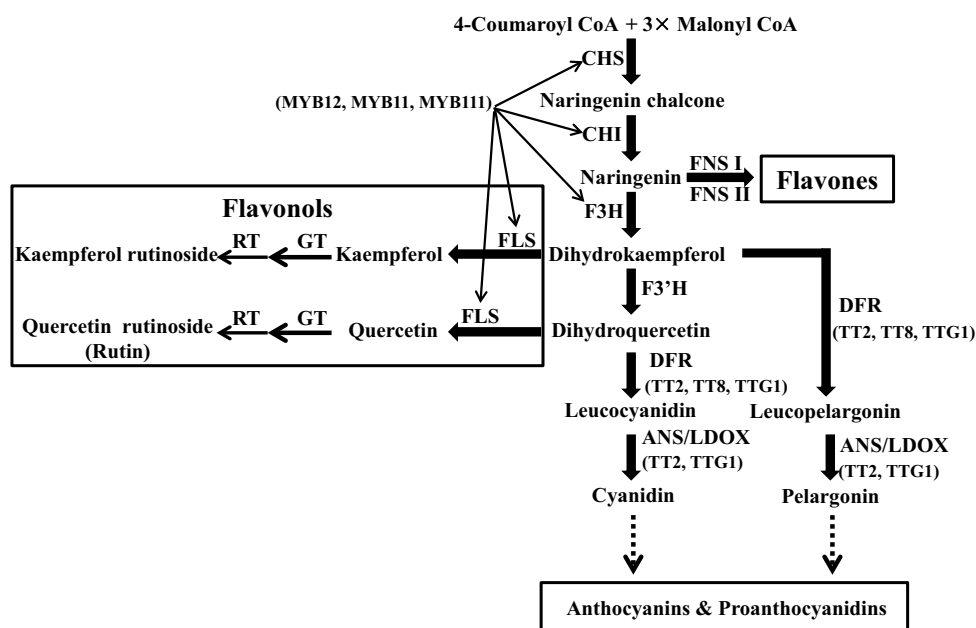
The major types of flavonoids in plants are flavonols, flavanones, flavones, anthocyanins, condensed tannins (proanthocyanidins), and so on (Routaboul et al. 2012). Flavonoids have antioxidant, anti-inflammatory, and anti-proliferative properties that can contribute to protecting humans from cancers and cardiovascular diseases (Geleijnse and Hollman 2008; Kale et al. 2008; Kaur et al. 2008; Kang et al. 2009; van Dam et al. 2013; Maggioni et al. 2014). Flavonoids are also known to participate in the regulation of the growth, development, and physiology of *Arabidopsis* and many other plants. For example, flavonols are involved in the modulation of the polar transport of auxin (Kuhn et al. 2011), flower color (Gronquist et al. 2001), pollen fertility (Mo et al. 1992), ethylene signaling (Lewis et al. 2011), and UV-B protection (Kusano et al. 2011). Thus, understanding and the promise of being able to engineer the synthesis of flavonoids, especially flavonols, is of great biological and medicinal significance.

The biosynthesis pathways of flavonoids in various plant species have been characterized through the isolation and characterization of numerous biosynthetic enzymes (Fig. 1) (Winkel-Shirley 2001; Williams and Grayer 2004; Kuhn et al. 2011; Routaboul et al. 2012; Cheng et al. 2014). In flavonoid biosynthesis, naringenin is hydroxylated to dihydroflavonol by F3H. Dihydroflavonol can subsequently be oxidized to flavonol by FLS (Chua et al. 2008; Owens et al. 2008). Alternatively, dihydroflavonol-4-reductase (DFR) reduces the dihydroflavonol to the leucoanthocyanidin, and then ANS, also known as LDOX/leucoanthocyanidin dioxygenase) produces an anthocyanin from the leucoanthocyanidin (Davies et al. 2003; Turnbull et al. 2004; Wellmann et al. 2006). F3H, FLS, and ANS all belong to the 2ODD

oxygenase family, and are widely distributed among various plant species. Plant 2ODD oxygenases utilize molecular oxygen as their co-substrate, and are distinguished by their differing co-factor requirements, including, for example, ascorbate, 2-oxoglutarate, and/or  $\text{Fe}^{2+}$  (Chua et al. 2008). These enzymes are involved in hypoxic signaling, DNA repair, and various metabolic pathways, including the biosynthesis of phytohormones, amino acids, signaling molecules, and various secondary metabolites (Prescott and Lloyd 2000; Ivan et al. 2001; Falnes et al. 2002; Treweek et al. 2002).

The first full-length *FLS* gene was cloned from a petunia petal cDNA library. Its activity was confirmed by antisense expression in petunia, which led to significant reductions of flavonol synthesis in petals (Holton et al. 1993). Subsequent studies have characterized *FLS* genes from a range of plant species, including *Arabidopsis thaliana* (Chua et al. 2008; Owens et al. 2008; Preuss et al. 2009; Stracke et al. 2009), *Vitis vinifera* (Fujita et al. 2006), *Camellia sinensis* (Lin et al. 2007), and *Zea mays* (Falcone Ferreyra et al. 2010). The first *ANS* gene was isolated from maize and a further investigation of its biochemical functions was performed in *Perilla frutescens* (Menssen et al. 1990; Saito et al. 1999). The crystal structure of ANS was determined and has been widely used as template for building *in silico* models of other plant 2ODD oxygenases (Welford et al. 2005). Such models are useful for the identification of conserved residues that are required for oxygenase activity (Welford et al. 2001). The *F3H* gene was first described in *Matthiola incana* (Forkmann et al. 1980), and subsequently cloned from *Petunia hybrid* (Britsch et al. 1992), *Hordeum vulgare* (Meldgaard 1992), *Malus* (Davies 1993), *Medicago sativa* (Charrier

**Fig. 1** Sketch map of the flavonoid biosynthetic pathway. *CHS* chalcone synthase, *CHI* chalcone isomerase, *F3H* flavanone-3-hydroxylase, *FNS I* flavone synthase I, *FNS II* flavone synthase II, *F3'H* flavonoid-3'-hydroxylase, *FLS* flavonol synthase, *GT* flavonol-3-glucosyltransferase, *RT* rhamnosyltransferase, *DFR* dihydroflavonol-4-reductase, *ANS* anthocyanidin synthase



et al. 1995), *Zea mays* (Deboo et al. 1995), *Arabidopsis thaliana* (Pelletier and Shirley 1996), *Triticum aestivum* (Himi et al. 2011), and soybean (Cheng et al. 2013). Recombinant domain swapping research found that the C-terminal region of F3H contributed to its activity but did not affect its selectivity (Lee et al. 2001; Wellmann et al. 2004). Although the 2ODD oxygenases have been extensively studied in a few plant species, such as *Arabidopsis*, maize, and grape, the 2ODD genes in tobacco have not been characterized.

In the current study, we identified and characterized the 2ODD genes involved in flavonoid biosynthesis from the allotetraploid *N. tabacum* and from two diploid tobacco species (*N. tomentosiformis* and *N. sylvestris*). Our analysis included the identification of tobacco 2ODD-IFB genes from genomic data, analysis of their structures and conserved motifs, phylogenetic analysis, evaluation of their N-/C-terminal structures, evaluation of their tissue- and developmentally dependent expression patterns, and correlations between 2ODD-IFB gene expression patterns and flavonoid content. Our results provide a foundation for the further investigation and manipulation of tobacco 2ODD-IFB genes, which should lead to a more complete understanding of the roles of these genes in the regulation of plant growth and may enable the intentional alteration of flavonoid composition and/or content according to human requirements.

## Materials and methods

### Plant materials

The Honghua Dajinyuan cultivar (Hongda, *N. tabacum*) was used in the expression profile analysis for the 2ODD-IFB genes. Tobacco seeds maintained at our laboratory were germinated and grown in pots under typical conditions (daily 28 °C for 16 h under light, 23 °C for 8 h of darkness) until the flowering stage. Total RNA was extracted from roots, leaves, stems, stem nodes, buds, sepals, stamens, pistils, and axillary buds.

For the stress treatments, tobacco seeds were soaked, sterilized, and then germinated on 1/2 MS medium in darkness. Young seedlings were then grown under typical culture conditions in vermiculite until the six-leaf stage. *Phytophthora parasitica* var. *nicotianae* (*P. parasitica* var. *nicotianae*) and potato virus Y (PVY) were inoculated onto the sixth leaves of tobacco seedlings for different durations of time (1 h, 2 h, 4 h for *P. parasitica* var. *nicotianae* inoculation; 3d, 6d, 9d, 14d for PVY inoculation). For the drought treatment, seedlings were initially transferred to a 1/3 strength Hoagland solution for 1 week prior to use in the experiment, and then cultivated in a solution containing 20% (w/v) PEG6000 for 1d, 2d, 4d, and 8d. Plant materials from the treatments were

collected and immediately frozen in liquid nitrogen and stored at –80 °C prior to RNA extraction.

### Phylogenetic and gene structure analyses

The 2ODD-IFB sequences involved in flavonoid biosynthesis in other plant species were collected from NCBI GenBank. The sequences used in this study are listed in Suppl. Table S1.

Multiple alignments of the 2ODD-IFB sequences were performed using Clustal X (version 1.83) and DNAMAN (version 6.0) with default penalties for gap and the protein weight matrix of Blosum (Jeanmougin et al. 1998). MEGA 5.0 was used to construct the phylogenetic tree of the 2ODD-IFB amino acid sequences using the neighbor-joining algorithm with bootstrap method (1000 of bootstrap replications), Poisson model, and complete deletion of gaps/missing data (Tamura et al. 2011).

Each tobacco 2ODD-IFB coding sequence was aligned with its corresponding genomic DNA sequence to identify the exact intron/exon locations. The Multiple Em for Motif Elicitation program (MEME, <http://meme.nbcr.net/meme3/mme.html>) was used to predict the conserved motifs in the tobacco 2ODD-IFB protein sequences (Bailey et al. 2006). The putative motifs found by MEME were further queried in the InterPro database (<http://www.ebi.ac.uk/interpro/jsearchionid=F412A1E32D81ECA4EBDB9A250D55D32E>) to explore their possible biological functions (Finn et al. 2017).

### cDNA sample preparation and gene transcription analysis

A SuperPure Plantpoly RNA Kit (Gene Answer, Beijing, China) was used to extract total RNA from the plant samples. DNA contamination was removed by digestion with RNase-free DNase I (Gene Answer). Reverse Transcriptase M-MLV (Takara Biomedical Technology, Beijing, China) and random primers were used to synthesize first-strand cDNA using 1 µg of total RNA as a template. The cDNA concentrations were evaluated with a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) and then diluted to a 100 ng/µl concentration.

The gene-specific primers used in the PCR experiment are listed in Suppl. Table S2. RT-PCR was performed using Takara Taq polymerase in a Biometra thermal cycler. The PCR program was as follows: 95 °C for 5 min, 25–29 cycles of 94 °C for 30 s, 55 °C or 60 °C for 30 s, and 72 °C for 30 s. qPCR amplification reactions were performed using a LightCycler® 96 SW 1.1 cycler (Roche, Laval, QC, Canada) and an SYBR Green kit (Roche). The PCR program was as follows: 95 °C for 5 min, 40 cycles of 94 °C for 30 s, 60 °C for 30 s, signal acquisition, and then a final melting curve of 65–95 °C. The expression levels of the target genes were

standardized to the expression level of the *NtGAPDH* gene using the  $2^{-\Delta\Delta Ct}$  method. Three independent biological replicates were performed for each gene.

### Estimation of $K_a/K_s$ ratios and protein structure modeling

To estimate  $K_a$  (nonsynonymous substitutions per site),  $K_s$  (synonymous substitutions per site) and their ratios, the amino acid sequences of each gene family were firstly analyzed by the MEME program. The motif amino acid sequences found by the program were aligned and subsequently transferred to the original cDNA sequences. Both  $K_a$  and  $K_s$  values were then estimated from their aligned cDNA sequences using the yn00 program of the PAML4b package.

The homology models of the tobacco 2ODD-IFB proteins were generated with tools available via the SWISS-MODEL website (<http://swissmodel.expasy.org/interactive>), as described in a previous paper (Guex et al. 2009). Each target protein sequence was aligned with those of model proteins to select the most suitable template. Based on various factors (insertions/deletions, ligands, oligomeric state) for which information was available on the SWISS-MODEL website, the crystal structure of *Arabidopsis* LDOX (2brt.1A) was chosen as the template from which to build models for the tobacco 2ODD-IFB proteins.

### Flavonoid staining

Tobacco seedlings (Hongda) grown in vermiculite medium were stained for 3 h in a solution containing saturated (0.25%, w/v) diphenylborinic anhydride (DPBA) and 0.2% (v/v) Triton X-100. The stained roots were visualized with an epifluorescence microscope (Olympus, Tokyo, Japan) equipped with an FITC filter (excitation 450–490 nm, suppression LP 515 nm).

### Virus-induced gene silencing (VIGS) of 2ODD-IFB in tobacco

For the VIGS vectors construction, fragments with about 500 bp length of tobacco 2ODD-IFB genes were selected and amplified from cDNA samples using gene-specific primers shown in Suppl. Table S2. The fragments were digested with BamH I and Kpn I enzymes, and subsequently cloned into the pTRV2 vector. After sequencing, correct pTRV2-FLS, pTRV2-ANS, and pTRV2-F3H vectors were transformed into *A. tumefaciens* strain GV3101 by heat shock methods. The *Agrobacterium* suspensions were incubated at room temperature for at least 3 h without shaking, and subsequently injected into the leaves of 4-week-old *N. benthamiana* plants using a needleless syringe. The *Agrobacterium* containing pTRV2, pTRV2-PDS vectors were set as negative

and positive control, respectively. The positive plants were identified by RT-qPCR 2 weeks after injection.

### Analysis of flavonoid contents in various tobacco tissues

The powder prepared from about 150 mg of fresh tobacco tissue samples was mixed with 1.5 ml 80% ethanol (containing 0.012 g/L vitexin as internal standard) and then extracted for 30 min with ultrasonication. The mix was centrifuged at 12,000g for 10 min at room temperature. The supernatant was then filtered through a 0.22  $\mu$ m membrane before being analyzed for flavonoid contents by HPLC–MS. The analytical method used was as follows: chromatographic column, ACQUITY UPLC<sup>®</sup>BEH Phenyl (1.7  $\mu$ m 2.1  $\times$  150 mm, Waters); gas temperature, 350  $^{\circ}$ C; injection volume, 1  $\mu$ L; flow, 0.3 mL/min; wave length, 230, 260, 360, and 570 nm; mobile phase, A 100% water, B 100% acetonitrile; gradient elution, 8 min 85% A + 15% B, 5 min 58% A + 42% B, 0.01 min 100% B, 3 min 95% A + 5% B, 18 min stop.

## Results

### Identification and sequence analysis of tobacco 2ODD-IFB genes

We used the terms ‘flavonol synthase’, ‘anthocyanidin synthase’, and ‘flavanone 3-hydroxylase’ as query key words for searching the *N. tabacum* genome in the China tobacco genome database v4.0 (data not shown) and obtained 7, 8, and 7 hits for these terms, respectively. We then performed several blastp searches in the database using the amino acid sequences of *Arabidopsis* FLS, ANS, and F3H proteins as the query sequences. We aligned the nucleotide and amino acid sequences of each putative tobacco 2ODD-IFB homolog with those of the *Arabidopsis* FLS, ANS, and F3H genes. We excluded the tobacco genes that had low similarity and identity with *Arabidopsis* homologs and thus identified two FLS, two ANS, and four F3H genes in the *N. tabacum* genome (Table 1). Similarly, we identified and characterized one FLS, one ANS, and one F3H gene in the *N. tomentosiformis* genome as well as one FLS, one ANS, and one F3H gene in the *N. sylvestris* genome (Table 1). All the 2ODD sequences identified in our database were further validated in the public tobacco genome databases (Sierra et al. 2013, 2014; Edwards et al. 2017), and then submitted to the GenBank database. The sequence information and GenBank accession numbers are shown in Suppl. Table S3.

In brief, we identified four FLS genes, four ANS genes, and six F3H genes from the three tobacco species. The identities of the four FLS amino acid sequences were between 66.04 and 94.51%. However, the identities between the

**Table 1** Characteristics of tobacco 2ODD genes involved in flavonoids biosynthesis

	Gene ID	Gene location	gDNA size (bp)	CDS size (bp)	Amino acid size	MW(kDa)	<i>pI</i>
<i>FLS</i>	Ntab 0125670	Ntab_scaffold_133	5649	1017	338	38.47	6.15
	Ntab 0215410	Chr05	5190	1041	346	39.15	6.43
	Ntom0342270	Ntom_superscaffold_42	5770	1017	338	31.2	6.8
	Nsyl0361410	Nsyl_superscaffold_192	5851	957	318	36.22	6.43
<i>ANS</i>	Ntab 0256350	Chr18	1512	1260	419	47.23	5.11
	Ntab 0694550	Ntab_scaffold_448	1524	1260	419	47.19	5.29
	Ntom0270640	Ntom_scaffold_512	1531	1260	419	32.75	5.12
	Nsyl0342440	Nsyl_superscaffold_313	1524	1260	419	47.19	5.29
<i>F3H</i>	Ntab 0333640	Chr08	1652	1110	369	41.37	6.2
	Ntab 0444890	Ntab_scaffold_2534	1675	1110	369	41.28	5.23
	Ntab 0908200	Ntab_scaffold_833	1675	1110	369	41.27	5.23
	Ntab 0908150	Ntab_scaffold_833	1675	1110	369	41.28	5.23
	Ntom0320920	Ntom_superscaffold_31	1652	1110	369	41.37	6.2
	Nsyl0326580	Nsyl_superscaffold_383	1676	1110	369	41.28	5.23

sequences of the four *FLS* and four *ANS* ranged from 24.05 to 32.14%, and those of the four *FLS* and the six *F3H* ranged from 17.19 to 25.4% (Table 2). The four *ANS* homologs from the three tobacco species shared identities between 63.01 and 100% and the six *F3H* homologs shared identities between 96.21 and 100%. The identities between *ANS* and *F3H* were very low (15.18–24.82%). Similarly, alignments between the coding sequences of the 2ODD-IFB genes showed consistent results with those for the amino acid sequences (Table 2). It is worth noting that the *NtF3H* genes Ntab 0444890, Ntab 0908200, and Ntab 0908150 in the *N. tabacum* genome showed 100% identities in the amino acid sequences and CDS sequences, but they were found to be located either on different scaffolds or on different positions of the same scaffold (Table 1). We thus considered these as three independent *NtF3H* genes, which might be generated by duplication after the formation of tetraploid tobacco.

### Exon/intron organization of the 2ODD-IFB genes

Characterization of the exon/intron structure of genes can be useful for the evaluation of the evolution of a gene family. As shown in Fig. 2a, all of the tobacco *FLS* genes contained two introns. Most of the known *FLS* genes from other eudicot plant species also contained two introns (Suppl. Fig. S1). Among all of the *FLS* genes examined in this study, the length of exon one changed to a greater extent (ranging from 380 to 518 bp) than did exon two (from 327 to 333 bp) or exon three (from 211 to 229 bp). It is interesting that the *FLS* intron one length changed in a manner consistent with the divergence time between certain plant species and *Arabidopsis* (Fig. 3). For instance, previous studies have shown that

the divergence of the Solanaceae and *Arabidopsis* occurred at about 112–156 Mya, a time earlier than the divergence of grape (*Vitis vinifera*, 101–112 Mya), cotton (*Gossypium arboreum*, 83–86 Mya), and *Brassica napus* (20–24 Mya) (Bowers et al. 2003; Moore et al. 2010). The *FLS* intron one in Solanaceae plants varies in length from 1.78 to 4.12 kb, lengths that are all much larger than those of intron one in grape (0.64 kb), cotton (0.31 kb), *Brassica napus* (0.25 kb), and *Arabidopsis* (0.21 kb). Although Brachypodium, maize, rice, and sorghum diverged from *Arabidopsis* even earlier (about 200 Mya), the *FLS* genes in these plants contained no or only very short introns at this position in the gene's structure (0.07 kb in *OsFLS*, and 0.09 kb in *SbFLS2*). We aligned the nucleotide sequences of *FLS* intron one from various plant species, including tobacco, potato, tomato, grape, and *Arabidopsis*. As shown in Suppl. Fig. S2, there were three conserved fragments among the different *FLS* intron one sequences. Fragment one shown in black frame was conserved among all the sequences examined, while fragment two shown in a red frame was conserved in the Solanaceae *FLS* genes. The conserved fragment three in purple frame only showed up in tobacco *FLS* genes. Similarly, intron two of the Solanaceae *FLS* genes was larger than those in other plant species, especially in the *StFLS* and *SolyFLS* genes (Fig. 3). Alignment among the intron two sequences from various plant species revealed two conserved fragments (Suppl. Fig. S3). The conserved fragment one existed in all the species examined, while the fragment two only occurred in the *StFLS* and *SolyFLS* genes.

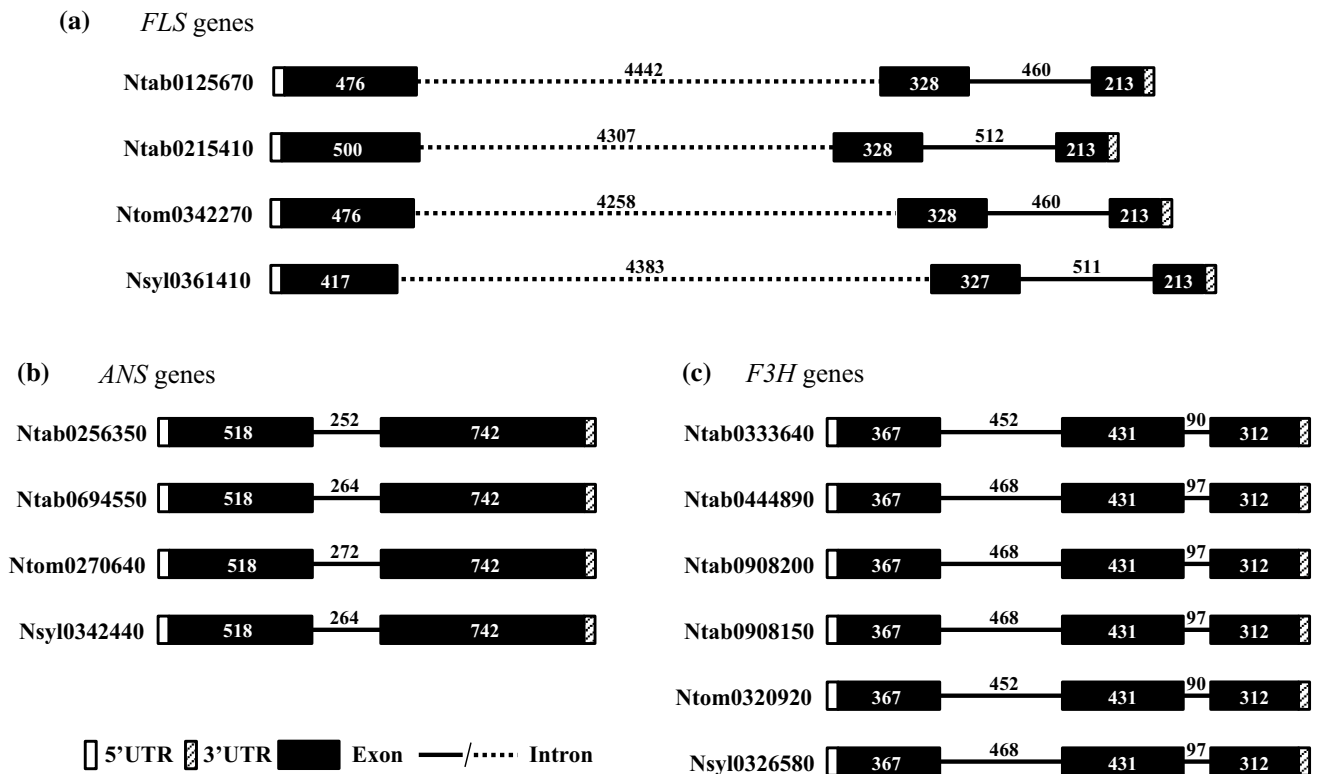
We also analyzed the structures of the *ANS* and *F3H* genes from various plant species. As shown in Figs. 2a, b and S4, all the *ANS* genes from eudicot plants examined, including tobacco, potato, tomato, *Arabidopsis*, *Brassica*



**Table 2** Identity matrix of tobacco ZODD-IPB amino acid sequences and their coding sequences

CDS identity		Amino acid identity											
		FLS				ANS				F3H			
Ntab	Nsy03	Ntab	Ntom03	Nsy03	Ntab	Ntom03	Nsy03	Ntab	Ntom03	Nsy03	Ntab	Ntom03	Nsy03
0125670	61410	0215410	42270	61410	0256350	0694550	0694550	0694550	0694550	0333640	0444890	0908200	0908150
		FLS				ANS				F3H			
FLS													
Ntab 0125670		94.51	80.18	91.12	29.52	30.95	22.81	30.95	22.81	22.99	22.46	22.46	22.46
Ntab 0215410	<b>93.08</b>		75.14	91.62	31.67	32.14	29.94	32.14	29.94	24.15	23.62	23.62	23.62
Ntom0342270	<b>70.89</b>	<b>67.15</b>		66.04	24.05	24.29	24.31	24.29	24.31	17.79	17.19	17.19	17.19
Nsy0361410	<b>89.77</b>	<b>91.74</b>	<b>72.83</b>		29.29	29.76	24.31	29.76	24.31	25.4	24.87	24.87	24.87
ANS													
Ntab 0256350	<b>41.3</b>	<b>43.2</b>	<b>33.2</b>	<b>39.48</b>		96.66	64.88	96.66	64.88	23.64	23.64	23.64	23.64
Ntab 0694550	<b>40.98</b>	<b>43.29</b>	<b>33.1</b>	<b>39.32</b>	<b>96.83</b>		63.01	100	63.01	24.82	24.82	24.82	24.82
Ntom0270640	<b>42.61</b>	<b>41.72</b>	<b>35.57</b>	<b>41.16</b>	<b>66.98</b>	<b>65.32</b>		63.01	63.01	15.45	15.18	15.18	15.18
Nsy0342440	<b>40.98</b>	<b>43.29</b>	<b>33.1</b>	<b>39.32</b>	<b>96.83</b>	<b>100</b>	<b>65.32</b>		65.32	24.82	24.82	24.82	24.82
Ntab 0333640	<b>43.46</b>	<b>42.26</b>	<b>33.87</b>	<b>40.28</b>	<b>42.03</b>	<b>42.82</b>	<b>36.26</b>	<b>42.82</b>	<b>36.26</b>	96.21	96.21	96.21	96.21
F3H													
Ntab 0444890	<b>42.44</b>	<b>41.04</b>	<b>32.89</b>	<b>39.49</b>	<b>39.95</b>	<b>40.84</b>	<b>36.29</b>	<b>40.84</b>	<b>36.29</b>	<b>95.41</b>		100	100
Ntab 0908200	<b>42.44</b>	<b>41.04</b>	<b>32.89</b>	<b>39.49</b>	<b>39.95</b>	<b>40.84</b>	<b>36.29</b>	<b>40.84</b>	<b>36.29</b>	<b>95.41</b>	<b>100</b>	100	100
Ntab 0908150	<b>42.44</b>	<b>41.04</b>	<b>32.89</b>	<b>39.49</b>	<b>39.95</b>	<b>40.84</b>	<b>36.29</b>	<b>40.84</b>	<b>36.29</b>	<b>95.41</b>	<b>100</b>	<b>100</b>	100
Ntom0320920	<b>43.46</b>	<b>42.26</b>	<b>33.87</b>	<b>40.28</b>	<b>42.03</b>	<b>42.82</b>	<b>36.26</b>	<b>42.82</b>	<b>36.26</b>	<b>100</b>	<b>95.41</b>	<b>95.41</b>	96.21
Nsy0326580	<b>42.44</b>	<b>41.04</b>	<b>32.89</b>	<b>39.49</b>	<b>39.95</b>	<b>40.84</b>	<b>36.29</b>	<b>40.84</b>	<b>36.29</b>	<b>95.41</b>	<b>100</b>	<b>100</b>	<b>95.41</b>

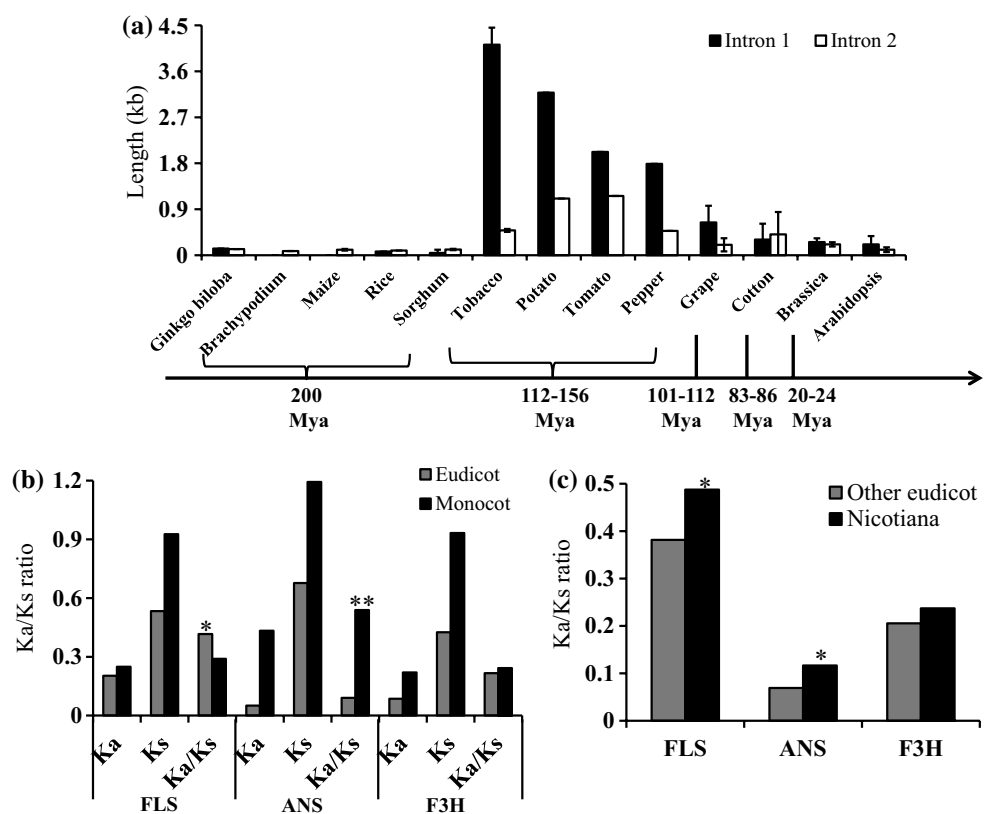
Italic represents the identities between different amino acid sequences, while bold represents the identities between different CDS sequences



**Fig. 2** Exon/intron structural organization of tobacco *FLS* **a**, *ANS* **b** and *F3H* **c** genes. Exons are denoted in black boxes, and the connecting lines represent introns. Numbers in boxes or above the lines represent the sizes (in bp) of the corresponding exons or introns, respectively. The 5' and 3' untranslated regions (UTRs) are represented by blank boxes

represent the sizes (in bp) of the corresponding exons or introns, respectively. The 5' and 3' untranslated regions (UTRs) are represented by blank boxes

**Fig. 3** Intron length and  $K_a/K_s$  values of 2ODD-IFB genes. **a** Intron length of the *FLS* genes from various plant species. The arrow below shows the predicted divergence time between the indicated plant species and *Arabidopsis*. **b** Values of  $K_a$ ,  $K_s$ , and  $K_a/K_s$  in monocot and eudicot plants. **c** Values of  $K_a/K_s$  in *Nicotiana* and other eudicot plants. The monocot and eudicot plant species are as shown in Supplementary Table S6



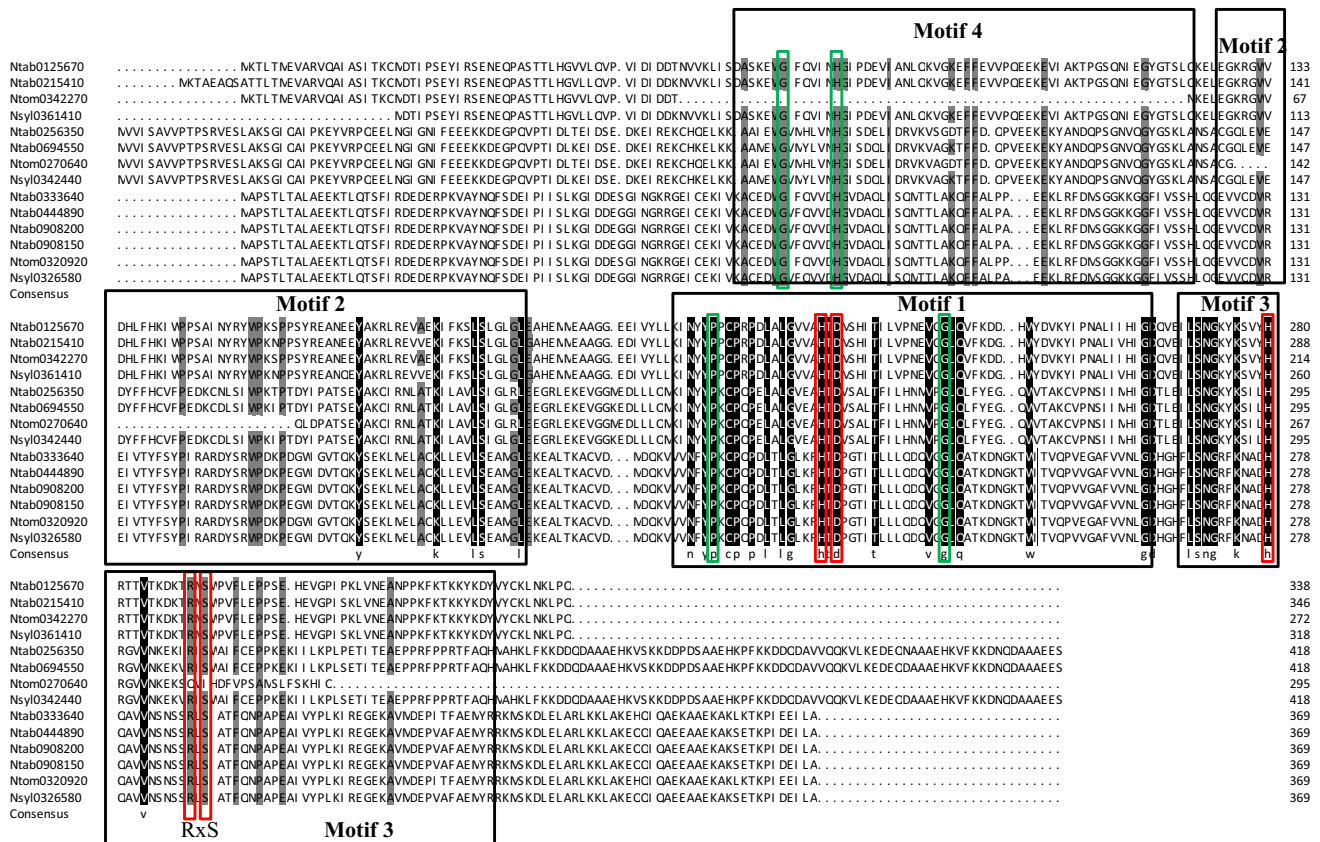
*napus*, soybean, *Medicago truncatula*, and grape, contained one intron and two exons. There was no intron present in the *ANS* genes of monocots including rice, maize, sorghum, or *Setaria italica*. The exon 1 size of the eudicot *ANS* genes varied from 497 to 533 bp, while those of exon two varied from 562 to 860 bp in length, suggesting a higher degree of conservation of exon one during the evolution of the *ANS* gene. All the *F3H* genes examined in the present study contained two introns, though the intron size varied among the different plant species (Suppl. Fig. S4). Moreover, the sizes of the three exons were highly conserved, especially exon two, which was 431 bp in the tobacco *F3H* genes, and 429 bp in the other *F3H* genes. This indicated a relatively evolutionarily stable gene structure for *F3H*.

### Conserved residues in tobacco 2ODD-IFB enzymes

To evaluate if the genes identified in tobacco were likely to encode functional 2ODD-IFB enzymes, we examined their conserved catalytic residues and motifs (Cheng et al. 2014). As shown in Fig. 4, there were five specific residues (in a red frame) that were highly conserved across all of the tobacco

2ODD-IFB proteins, including two histidines (His224, His280; numbers refer to the Ntab 0125670 sequence), one aspartic acid (Asp226), one arginine (Arg290), and one serine (Ser292). Four other residues, Gly70, His77, Gly240, and Pro210 (shown in a green box) that are likely involved in ensuring the correct folding of 2ODD proteins, were also found to be conserved in the selected tobacco 2ODD-IFB proteins. We searched all of the putative motifs of the tobacco 2ODD-IFB proteins on the MEME website and found a total of four conserved motifs (Fig. 4, Suppl. Table S4). These motifs were all either annotated as possessing oxidoreductase activity or as being involved in anthocyanidin biosynthesis. Taken together, all of the tobacco 2ODD-IFB identified from the database contained the conserved residues and motifs that are considered characteristic of 2ODD proteins.

To identify the most likely enzymatic activity-relevant residues in the termini of the tobacco 2ODD-IFB proteins, we constructed homology models based on the crystal structure of AtLDOX (2brt.1.A) as a template. The tobacco NtFLS (Ntab 0125670), NtANS (Ntab 0694550), and NtF3H (Ntab 0333640) proteins exhibited 43.38%, 74.71%, and



**Fig. 4** Alignment of the predicted amino acid sequences of the tobacco 2ODD-IFB proteins. The red boxes indicate conserved catalytic residues in these 2ODD-IFB sequences, while the green boxes

indicate an additional four conserved residues that are probably required to ensure correct folding. The black boxes indicate the location of the putative motifs shown in Table S2

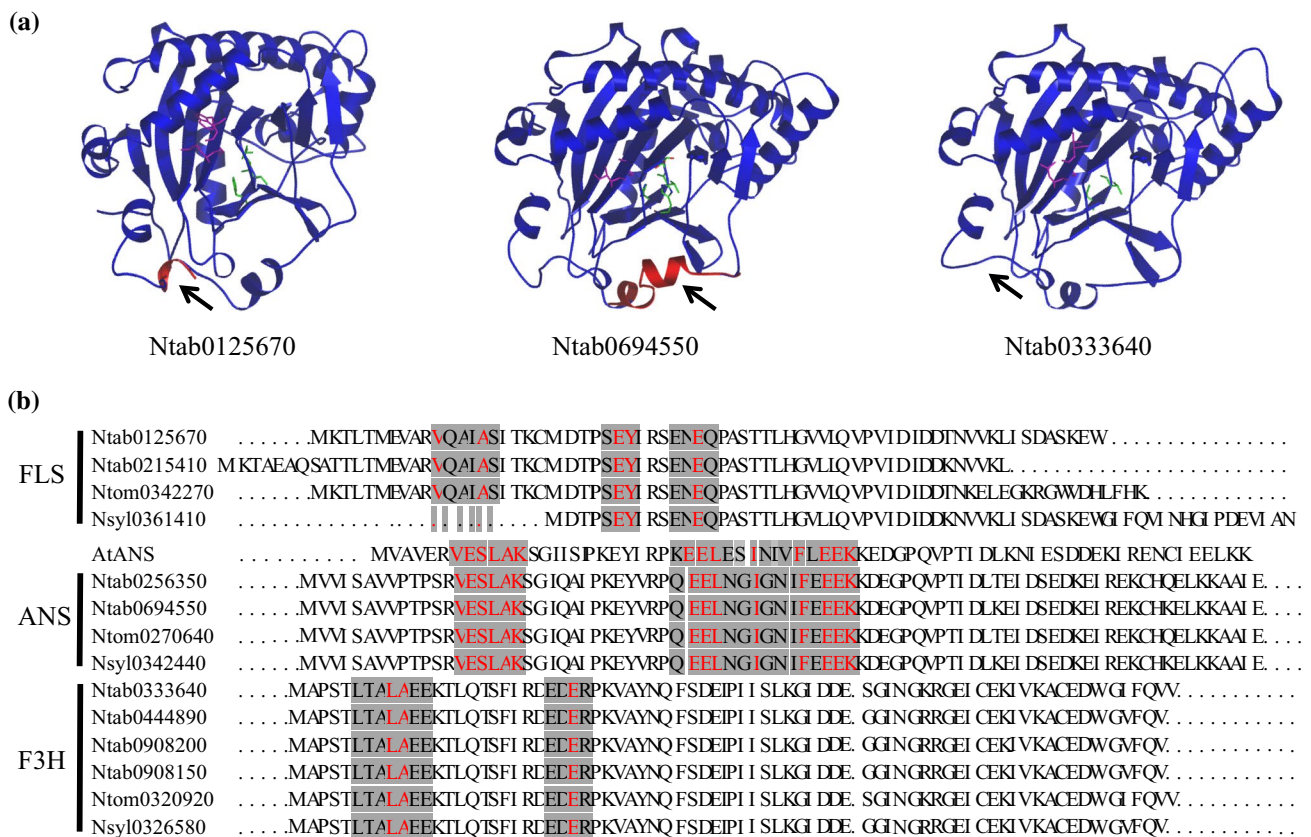


30.40% amino acid identity with AtLDOX, respectively. The largest apparent differences among the three protein structures were near the N terminus (Fig. 5a). There were at least two  $\alpha$ -helices at the N terminus of all three proteins. There was one more four-residue  $\alpha$ -helix at the N terminus of the NtANS model, while the NtFLS model contained half of an  $\alpha$ -helix at a similar position. This helix was not present in the NtF3H model. We also used the homology models to identify the conserved residues that might be required for the formation of the N-terminal helices. As shown in Fig. 5b, residues highlighted in gray were supposed to constitute the helices at the N terminus of the tobacco 2ODD-IFB proteins. The residues presented in red did not differ from those of the template and might be the core position for N-terminal helices.

**Phylogenetic analysis of tobacco 2ODD-IFB genes**

To better understand the evolutionary relationships among the 2ODD-IFB genes of tobacco and those of other plant species, we aligned 75 2ODD-IFB amino acid sequences

from 26 species with the ClustalX program and then constructed an unrooted tree with MEGA 5 using the neighbor-joining method. As shown in Fig. 6, plant 2ODD-IFB genes could be generally divided into two major groups. One group contained two branches, namely an FLS group and an ANS group, while the other group was divided into an F3H group and an FNS I group. The FLS group could further be divided into three subgroups, which we named as the FLS monocot subgroup, the FLS eudicot I subgroup, and the FLS eudicot II subgroup. There were 24 genes from 11 plant species in the FLS eudicot I subgroup, but only three genes (*GbFLS*, *VvFLS5*, and *CmFLS*) present in the FLS eudicot II subgroup. Moreover, in both the ANS group and the F3H group, the monocot genes were apparently separate from the eudicot genes. Interestingly, we found that PhANS (BAA89008), BnFLS3 (CDX91472), and BnANS (CDX68770) were grouped into the FNS I group, which was confined to rice and Apiaceae species in previous studies (Britsch et al. 1981; Prescott et al. 2002). This might indicate a more broad existence of *FNS I* genes in plants.



**Fig. 5** Structural analysis of the tobacco 2ODD-IFB proteins. **a** Homology models of NtFLS (Ntab 0125670), NtANS (Ntab 0694550), and NtF3H (Ntab 0333640) generated based on the crystal structure of AtANS (At4g22880). The arrows indicate the different helix motifs present in the N terminus of the tobacco 2ODD-IFB

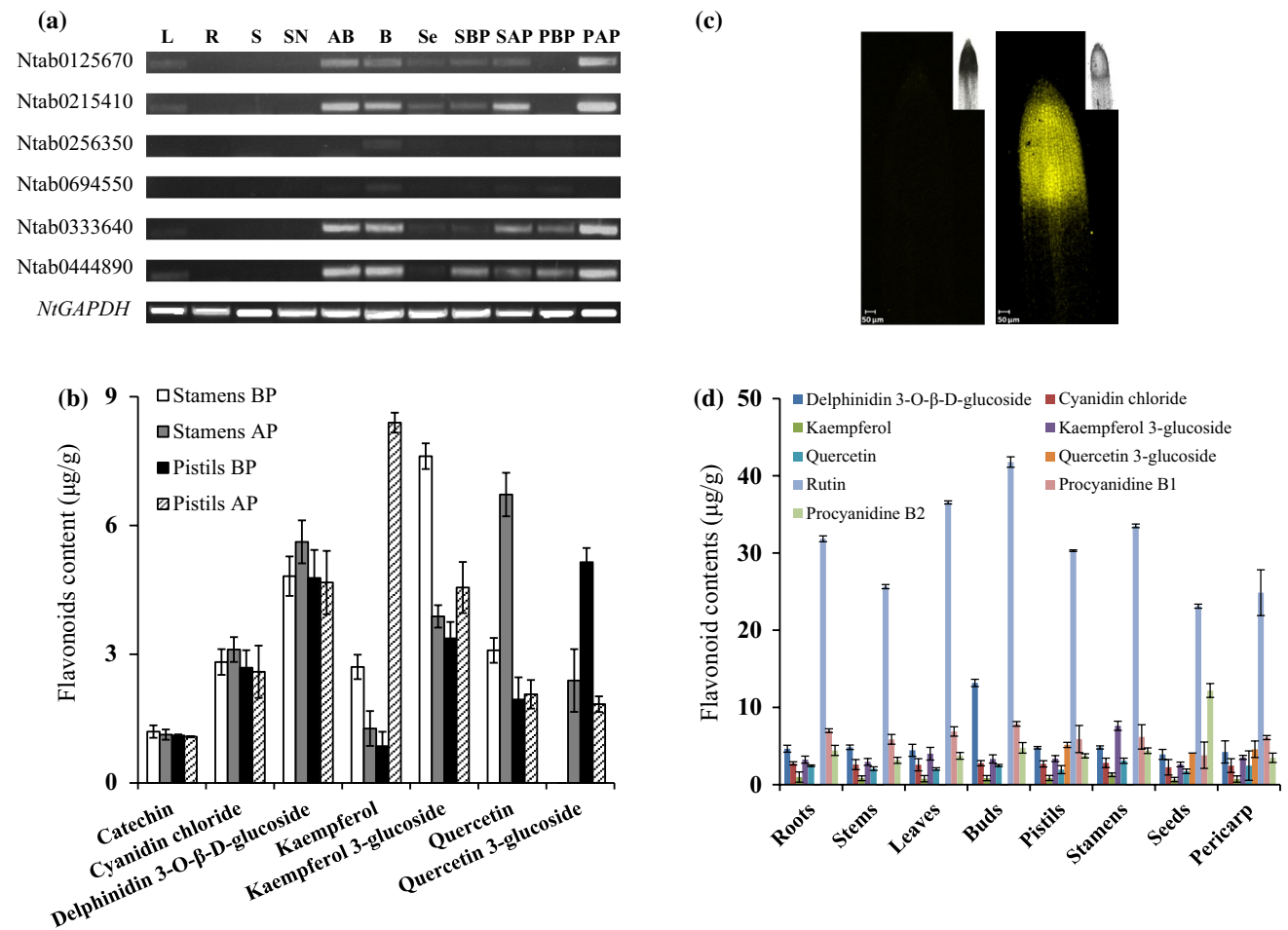
proteins. **b** Alignment of the 2ODD-IFB N-terminal sequences. The regions highlighted in gray are predicted to form the helix motifs at the N terminus. The residues shown in red are strictly conserved between the tobacco sequences and the AtANS sequence



### Correlation between 2ODD-IFB expression levels and flavonoids accumulation patterns

To better understand the possible functions of the 2ODD-IFB genes in tobacco, we initially examined their tissue-specific expression patterns in roots, leaves, stems, stem nodes, axillary buds, buds, sepals, stamens (before and after pollination), and pistils (before and after pollination). As shown in Fig. 7a, the expression of *NtFLS* (Ntab 0125670 and Ntab 0215410) was detected in leaves, buds, sepals, stamens, and pistils. The *NtANS* gene (Ntab 0256350 and Ntab 0694550) was mainly expressed in buds and sepals, while expression of *NtF3H* (Ntab 0333640 and Ntab 0444890) was detected in flower tissues, including buds, sepals, stamens, and pistils (Fig. 7a). Interestingly, the expression level of *NtFLS* changed markedly in pistils before and after pollination. The

expression level of *NtFLS* could hardly be detected in the pistils before pollination, but increased substantially after pollination. We then measured the content of kaempferol, kaempferol 3-glucoside (K-3G), quercetin, and quercetin 3-glucoside (Q-3G), four products of the FLS enzyme, in the stamens and pistils. As shown in Fig. 7b, pollination could lead to a strong decrease in the kaempferol and K-3G content in the stamens, but led to markedly increased kaempferol and K-3G content in the pistils. Moreover, the quercetin and Q-3G content in the stamens increased significantly after pollination, while the content of Q-3G was reduced in the pistils after pollination. Pollination did not alter the content of cyanidin chloride or delphinidin 3-*O*- $\beta$ -D-glucoside, two important products of *NtANS*, in the stamens or pistils. Thus, the variation in 2ODD-IFB gene expression in the stamens and pistils appears to be related to corresponding



**Fig. 7** Expression levels of the 2ODD-IFB genes and flavonoids content in different organs of *N. tabacum*. **a** RT-PCR analysis of tobacco 2ODD-IFB genes. RNA samples were collected from roots (R), leaves (L), stems (S), stem nodes (SN), axillary buds (AB), buds (B), sepals (Se), stamen before pollination (SBP), stamen after pollination (SAP), pistils before pollination (PBP), and pistils after pollination (PAP). The PCR amplifications for each gene were performed with

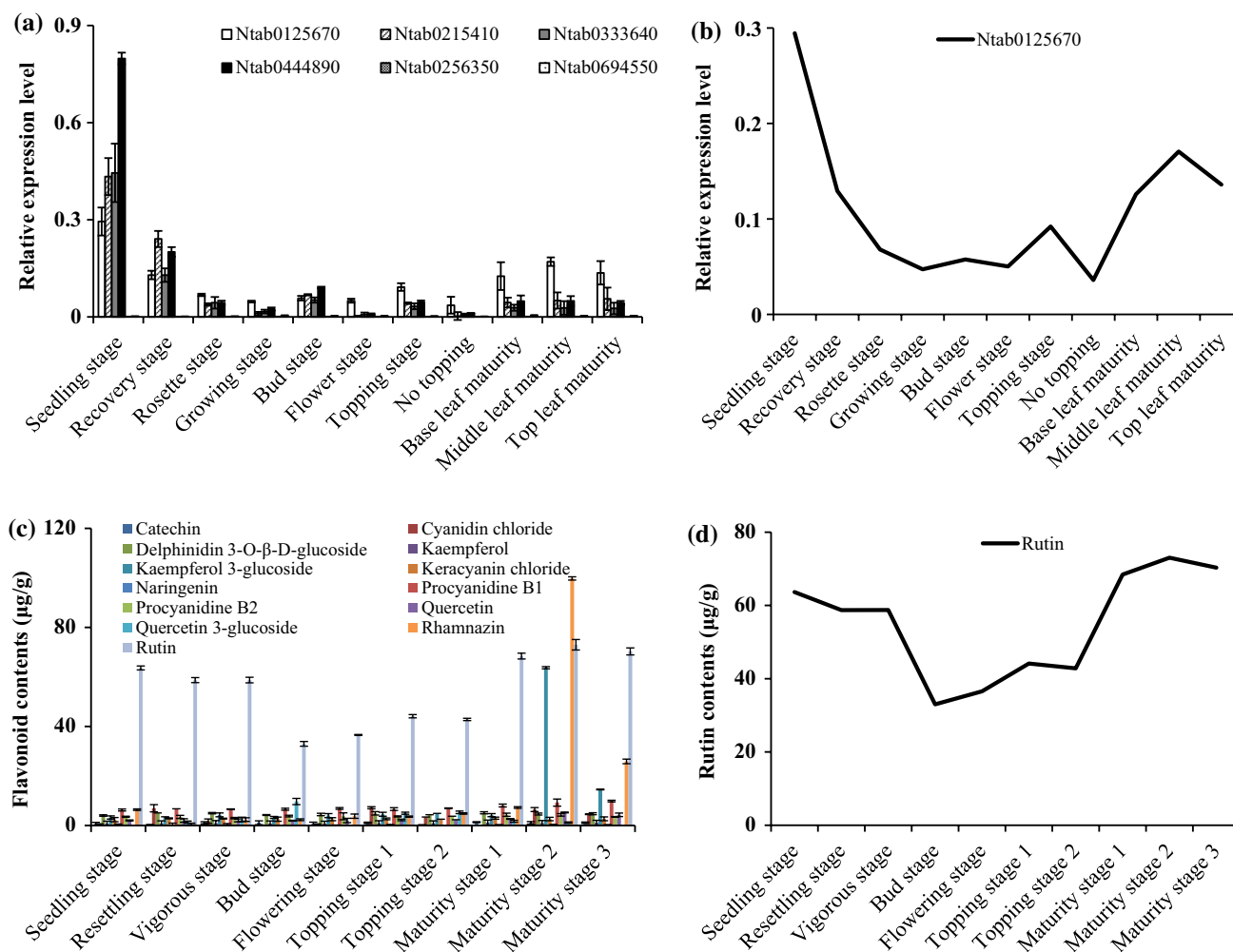
the same program, and the gels were run under the same experimental conditions. The full-length gels are shown as Suppl. Fig. S7. **b** Contents of seven flavonoids in tobacco flowers before and after pollination. BP, before pollination; AP, after pollination. **c** DPBA staining of flavonoids in root tips. Left, stained with water; right, stained with DPBA. **d** Contents of 12 flavonoids in different organs of *N. tabacum*

changes in flavonol content, which might further affect the development of stamens and pistils.

We observed the distribution of flavonoids in tobacco roots via DPBA staining. As shown in Fig. 7c, tobacco root tips were stained to deep yellow, indicating an abundance of naringenin chalcone or quercetin in these regions. In contrast, there was weak yellow staining in the elongation zone (not shown), indicating an asymmetrical distribution of flavonoids in tobacco roots. We analyzed the accumulation of nine flavonoids in various tobacco tissues (Fig. 7d) and found that almost all of the flavonoids examined could be detected in all of the tissues, a finding that was not consistent with the expression patterns of the corresponding 2ODD-IFB genes. For instance, the transcription of 2ODD-IFB genes could hardly be detected in tobacco roots, but the

accumulation of various flavonoids was found in roots via both DPBA staining and HPLC–UV analysis. The accumulation levels of most flavonoids we assessed were quite low (less than 5  $\mu\text{g/g}$  fresh weight), which might be caused by the fact that most substances we examined were the intermediate products during flavonoids metabolism.

Quantitative real-time RT-PCR was performed to further evaluate the relative expression levels of the 2ODD-IFB genes in tobacco leaves at different developmental stages (Fig. 8). Similar to the results of the semi-quantitative RT-PCR analysis, the expression of *NtANS* was hardly detected in all the tissues. *NtF3H* was expressed in leaves at all of the stages, but at varying levels (Fig. 8a). The *NtFLS* gene was apparently expressed in all of the leaf stages examined. The expression level of *NtFLS* appeared as a ‘U’ model across



**Fig. 8** Transcription levels of 2ODD-IFB genes and flavonoid contents in *N. tabacum* leaves at various developmental stages. **a** Relative expression levels of 2ODD-IFB genes in *N. tabacum* leaves. **b** The ‘U’ model expression pattern of the Ntab 0125670 gene during the course of development. The expression levels of the 2ODD-IFB genes were determined relative to the internal control gene,

*NtGAPDH*. Values are mean  $\pm$  SD of three independent biological replicates. **c** Flavonoid contents in *N. tabacum* leaves at various developmental stages. **d** The variation curve of rutin content in *N. tabacum* leaves at various developmental stages. Values are mean  $\pm$  SD of three independent biological replicates



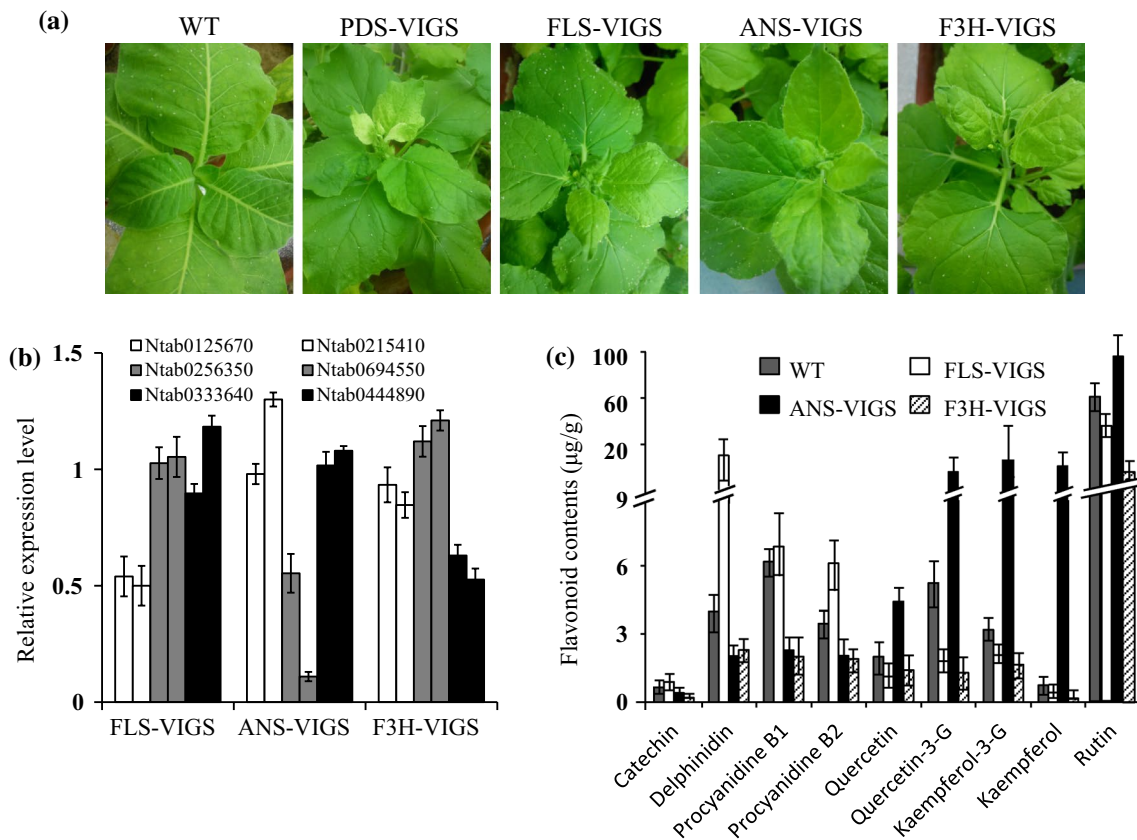
the whole developmental period (Fig. 8b), suggesting positive roles for this gene in both young and senescent leaves. It is worth pointing out that topping, an important agronomical practice in tobacco production, strongly induced the transcription of the *NtFLS* and *NtF3H* genes. We analyzed the flavonoid content in tobacco leaves at various developmental stages using HPLC–UV, and found that all of the leaves examined contained high levels of rutin (quercetin rutinoside) (Fig. 8c). The variation of the rutin content in leaves from different developmental stages also showed the aforementioned ‘U’ pattern (Fig. 8d), which was the same with the expression pattern of the *NtFLS* gene.

The tobacco 2ODD-IFB gene was silenced via the VIGS method to explore the relationship between flavonoids contents and 2ODD-IFB gene expression levels. As shown in Fig. 9a, b, the expression levels of the target genes were obviously down-regulated, while the phenotype of the VIGS plants was the same as that of the wild-type plants. The contents of anthocyanidins including catechin, delphinidin, and procyanidine in FLS–VIGS leaves were much more than those in wild-type leaves, but the flavonol contents including quercetin, quercetin-3-glucose, kaempferol,

kaempferol-3-glucose in FLS–VIGS leaves were less than those in WT leaves (Fig. 9c). On the contrary, compared to those in WT leaves, the anthocyanidins contents in the ANS–VIGS leaves decreased, but the flavonols contents significantly increased. The contents of flavonols and anthocyanidins both decreased in F3H–VIGS leaves. These data suggested that the contents of flavonoids in tobacco leaves were consistent with the expression levels of 2ODD-IFB genes.

### Differential expression patterns of tobacco 2ODD-IFB genes under drought and virus inoculation treatments

To examine any possible roles for the 2ODD-IFB genes in plant stress resistance, we measured their expression levels under drought and virus inoculation treatments by both qPCR (Fig. 10). Generally, the expression levels of the tobacco 2ODD-IFB genes were down-regulated, to varying extents, in leaves inoculated with *Phytophthora parasitica* var. *nicotianae* (*P. parasitica* var. *nicotianae*) for 1 h, 2 h, and 4 h (Fig. 10a). The variation trends of the expression levels of *NtFLS*, *NtANS*, and *NtF3H* genes were similar to each

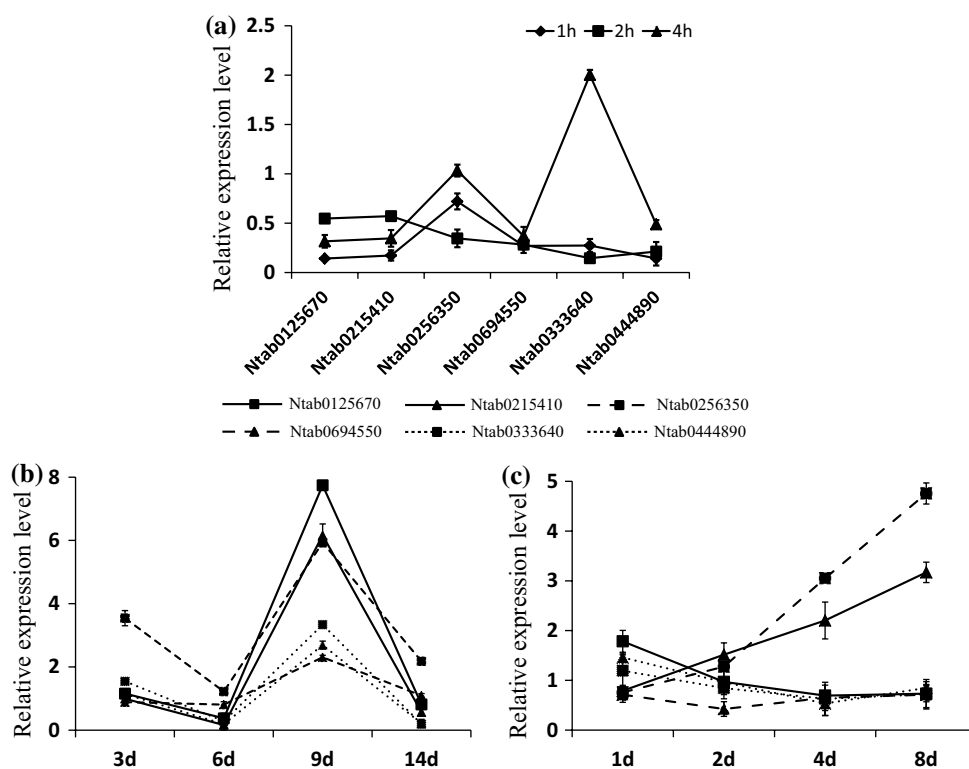


**Fig. 9** VIGS of 2ODD-IFB genes altered the flavonoids composition in tobacco leaves. **a** Phenotype of the VIGS tobacco plants. **b** Expression levels of tobacco 2ODD-IFB genes in VIGS plants. **c** The flavo-

noids contents in VIGS leaves. Values are mean ± SD of three independent biological replicates



**Fig. 10** Relative expression levels of the 2ODD-IFB genes in response to *P. parasitica* var. *nicotianae* inoculation **a**, potato virus Y (PVY) inoculation **b**, and drought treatment **c**. Values are mean  $\pm$  SD of three independent biological replicates



other in leaves inoculated with potato virus Y (PVY). As shown in Fig. 10b, compared with the expression levels in control leaves, the expression levels of the tobacco 2ODD-IFB genes were almost unchanged in leaves inoculated for 3 days (except a 3.5-fold increase in expression for the Ntab 0256350 gene), obviously suppressed in 6d leaves, then significantly up-regulated in 9d leaves, and finally decreased to various levels in 14-d-old leaves. When suffering drought stress, the expression patterns of the tobacco 2ODD-IFB genes were divided into two types (Fig. 10c). The expression levels of Ntab 0215410 (*NtFLS*) and Ntab 0256350 (*NtANS*) significantly increased along with the extension of the treatment time, but those of the other 2ODD-IFB genes were affected to a lesser extent throughout the treatment. Taken together, the expression levels of tobacco 2ODD-IFB genes under biotic stresses (*P. parasitica* var. *nicotianae* or PVY) fluctuated drastically along with the treatment time, while those under drought treatment showed apparent regularity, suggesting a more complex regulation network of tobacco 2ODD-IFB genes under biotic stresses relative to the regulation network for drought.

## Discussion

The 2ODD enzymes constitute the second largest protein superfamily in plants and participate in various oxygenation/hydroxylation reactions. Here, we identified four *FLS* genes,

four *ANS* genes, and six *F3H* genes from the *N. tabacum*, *N. sylvestris*, and *N. tomentosiformis* genomes. We further analyzed their conserved residues, gene structures, evolutionary relationships, and expression patterns in various tobacco plant materials. This work therefore provides an important step toward a comprehensive understanding of the phylogenesis and putative functions of these enzymes in various growth and developmental processes in tobacco.

## Evolution and diversity of tobacco 2ODD-IFB genes

Synthesizing the structural information we have learned from tobacco and other plant species, we proposed the events occurred in the structures of 2ODD-IFB genes during evolution (Fig. 11). Initially, the ancestral gene of 2ODD-IFB contained three exons and two introns. Before the split of gymnosperms and angiosperms, this ancestral gene gave rise to the three progenitors of *FLS*, *ANS*, and *F3H* genes via mutations or duplication. After the separation of the two kinds of plants, the angiosperm *ANS* genes lost their second intron, which led to the fusion of exon two and exon three (Suppl. Fig. S4). Subsequently, after the divergence of monocots and eudicots, the monocot *ANS* genes further lost their first intron, which resulted in no intron being present in these genes. Similarly, the first intron was also lost during the evolution of monocot *FLS* genes after the separation of monocots and eudicots (Suppl. Fig. S1). By contrast, the gene structures of *F3H* were quite stable among various

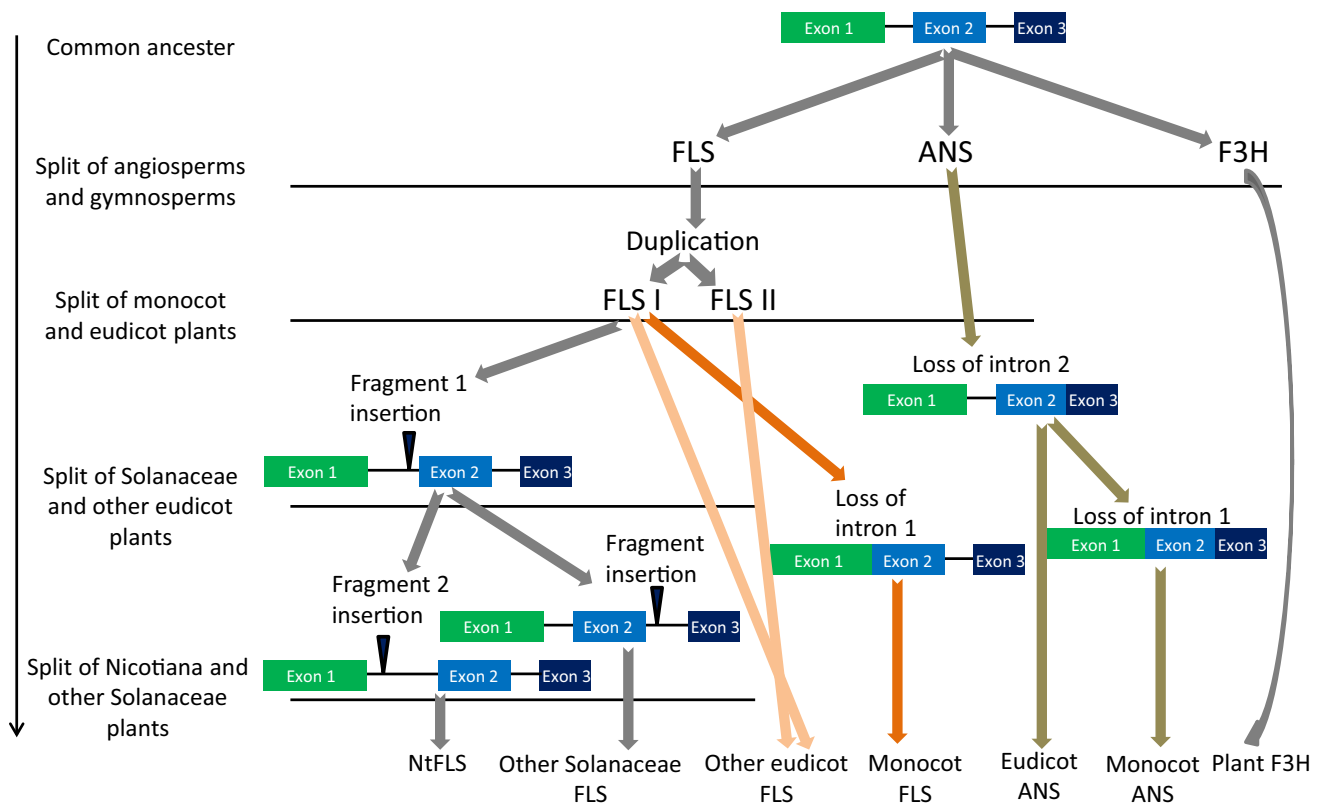


Fig. 11 Gene structure evolution of plant 2ODD-IFB genes

plant species examined, suggesting a relatively small pressure on this gene during evolution.

Besides the inconsistent numbers of introns and exons in plant 2ODD-IFB genes, we also found that the intron length of *FLS* genes varied greatly among various plant species (Fig. 3). Based on the alignment of different plant intron sequences, it is easy to come to the speculation that there was a fragment insertion (1st insertion) in Solanaceae *FLS* intron one position after its divergence from other plant species, but before the differentiation within the Solanaceae. Moreover, another fragment insertion (2nd insertion) occurred in the tobacco *FLS* intron one, which gave rise to the largest intron one examined so far. We also aligned the *FLS* intron two sequences from different plant species, and found a fragment insertion (3rd insertion) occurred in tomato and potato *FLS* genes (Suppl. Fig. S3). It has been hypothesized that the intron size is constrained by energy use in transcription, since it might require more energy to transcribe and splice larger introns (Castillo-Davis et al. 2002). So selection usually acts against introns with excessive size, but meanwhile there also exists a selective preference for larger introns in certain genes, owing to the possession of more regulatory elements and better control of gene transcription (Marais et al. 2005). The insertions of transposable elements (TE) in introns could result in significant genetic and phenotypic

changes via altering the spatial and temporal expression patterns of specific genes, which could also be preserved under natural selection (Lempe et al. 2005).

We have made several blast searches of the inserted sequences in *FLS* introns in the NCBI database, but failed to get highly similar sequences. There were many predicted TEs presented in the inserted fragments (Suppl. Table S5–7), indicating a more complex regulation system for the *FLS* genes harboring larger introns. The expression patterns of *FLS* genes vary among different plant species. For instance, the highest transcript levels of the two *FLS* genes in maize were found in young leaves (Falcone Ferreyra et al. 2012), but in *G. biloba* the highest *GbFLS* mRNA level was observed in the mature leaves and significantly higher than in the young leaves (Xu et al. 2012). In *Arabidopsis*, the highest transcript levels of *AtFLS* gene were detected in the reproductive tissues, including developing inflorescence, buds, flowers, and siliques (Owens et al. 2008). We found the tobacco *NtFLS* gene was mainly expressed in flowers and leaves. The various intron lengths might contribute to the distinct expression patterns for *FLS* genes in different plant species, which needed to be further confirmed.

The expansion of the 2ODD family occurred by large-scale duplication after the split from the common ancestor of land plants, which gave rise to three classes (DOXA, DOXB,

and DOXC) based on the amino acid sequence similarity. All the 2ODD-IFB belong to the DOXC class, and the evolution of FLS and ANS is likely to have occurred after the F3H emergence during seed plant evolution (Kawai et al. 2014). The 2ODD-IFB genes (*GbFLS*, *GbANS*, and *GnF3H*) of *Ginkgo biloba* dispersed in the FLS, ANS, and F3H group, respectively, instead of clustering together. This implied that the 2ODD-IFB genes might split apart from each other before the divergence of gymnosperms and angiosperms, which was consistent with the previous speculation in this text conducted by the analysis of gene structures. Moreover, we found that the FLS group could be further divided into three subgroups, among which one subgroup contained *FLS* genes from monocots, while genes in the other two subgroups were both from eudicots (Fig. 6). This suggests that duplication might have happened to the ancestral gene and given rise to the two progenitors of *FLS* genes before the divergence of monocots and eudicots. Subsequently, the two progenitors underwent independent evolution, and one progenitor generated the FLS eudicot I subgroup and FLS monocot subgroup, while the other one generated the FLS eudicot II subgroup. If so, there should be an FLS monocot II subgroup, the possibility of which we could not rule out, as there have been relatively fewer *FLS* genes reported in the monocot plants so far.

### Consistent 2ODD-IFB gene expression levels and flavonoids content in tobacco

In the present study, we measured the transcription levels of tobacco 2ODD-IFB genes in different tissues from flowering stage and found that most tobacco 2ODD-IFB genes were exclusively expressed in the reproductive tissues, including buds, sepals, stamens, and pistils. However, the deposition pattern of flavonoids among various tobacco tissues seemed not consistent with the expression patterns of 2ODD-IFB genes. For instance, we found flavonoids accumulation in tobacco roots by both DPBA staining and HPLC–UV analysis (Fig. 7c, d), while at the same time we did not detect obvious transcripts of tobacco 2ODD-IFB genes in this tissue (Fig. 7). This indicated the possibility that the flavonoids deposited in roots might be synthesized in other plant tissues, such as leaves, and then transported into the roots.

It is worth noting that in the tobacco tissues, where flavonoids were biosynthesized, the accumulation patterns of flavonoids were consistent with the expression patterns of 2ODD-IFB genes. For instance, pollination did not change the expression level of *NtANS* gene in the pistils, but significantly raised the transcription level of *NtFLS* gene (Fig. 7a). This led to the significant increase of flavonol content (especially, kaempferol) in pistils after pollination (Fig. 7b), which meanwhile also indicated a

positive role of flavonol in promoting pollen germination or pollen tube elongation. In tobacco leaves, both the expression pattern of *NtFLS* gene and the accumulation pattern of rutin turned to be a ‘U’ model during leaf development (Fig. 8b, 9d). Moreover, when *NtFLS* was silenced in tobacco leaves by VIGS, the biosynthesis of flavonoids flowed to the anthocyanidin pathway (Fig. 9c). More flavonols accumulated in the tobacco leaves where the *NtANS* gene was silenced. Taken together, it could be concluded that the expression patterns of tobacco 2ODD-IFB genes did not determine the deposition patterns of flavonoids among various tobacco tissues, but strongly affected the levels of flavonoids contents in the tissues, where they were biosynthesized.

### Distinct expression files of tobacco 2ODD-IFB genes under stresses

Flavonoids have been shown to participate in plant resistance to various biotic and abiotic stresses (Treutter 2005). In the present study, we detected the expression patterns of tobacco 2ODD-IFB genes under various stresses, and found two interesting facts. Firstly, the expression patterns of these genes varied under different biotic stresses. These distinct expression patterns suggested different roles played by 2ODD-IFB genes or flavonoids compounds in plant resistance to various biotic stresses. Secondly, we found that the expression patterns of up-regulated 2ODD-IFB genes were different among the plants with biotic and abiotic treatment, respectively. The up-regulation of 2ODD-IFB genes was transient under biotic stresses, while the up-regulation of these genes lasted longer under abiotic stress. We were inclined to believe that the regulation of the expression patterns of tobacco 2ODD-IFB genes might be in different ways under the two kinds of stresses, while those under biotic stresses were more complicated.

**Author contribution statement** JY conceived and supervised this study. ZW carried out most of the work. ZFL and SSW carried out the bioinformatics analysis. PPL determined the contents of flavonoids in tobacco tissues. QSC performed the microscopic examination and photograph of flavonoids staining. MZW and FL examined the gene expression levels and carried out the stress and VIGS treatments. JY and ZW summarized the results and wrote the manuscript. AGY advised and modified the manuscript. All authors reviewed the manuscript.

**Acknowledgements** This work was supported by the Project of Tobacco Transcription Factors Array Design and Expression Regulation of Key Transcription Factors (902014AA0520), the Project of

Regulation of the Transcription Factor NtMYB27 in the Biosynthesis of Flavonoids (902017CA0220), and the National Natural Science Foundation of China (902017AZ0340).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34:W369–373. <https://doi.org/10.1093/nar/gkl198> (Web Server issue)
- Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422(6930):433–438. <https://doi.org/10.1038/nature01521>
- Britsch L, Heller W, Grisebach H (1981) Conversion of flavanone to flavone, dihydroflavonol and flavonol with an enzyme system from cell culture of parsley. *Z Naturforsch C* 36:742–750
- Britsch L, Ruhnau-Brich B, Forkmann G (1992) Molecular cloning, sequence analysis, and in vitro expression of flavanone 3 beta-hydroxylase from *Petunia hybrida*. *J Biol Chem* 267(8):5380–5387
- Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA (2002) Selection for short introns in highly expressed genes. *Nat Genet* 31(4):415–418. <https://doi.org/10.1038/ng940>
- Charrier B, Coronado C, Kondorosi A, Ratet P (1995) Molecular characterization and expression of alfalfa (*Medicago sativa* L.) flavanone-3-hydroxylase and dihydroflavonol-4-reductase encoding genes. *Plant Mol Biol* 29(4):773–786
- Cheng H, Wang J, Chu S, Yan HL, Yu D (2013) Diversifying selection on flavanone 3-hydroxylase and isoflavone synthase genes in cultivated soybean and its wild progenitors. *PLoS One* 8(1):e54154. <https://doi.org/10.1371/journal.pone.0054154>
- Cheng AX, Han XJ, Wu YF, Lou HX (2014) The function and catalysis of 2-oxoglutarate-dependent oxygenases involved in plant flavonoid biosynthesis. *Int J Mol Sci* 15(1):1080–1095. <https://doi.org/10.3390/ijms15011080>
- Chua CS, Biermann D, Goo KS, Sim TS (2008) Elucidation of active site residues of *Arabidopsis thaliana* flavonol synthase provides a molecular platform for engineering flavonols. *Phytochemistry* 69(1):66–75. <https://doi.org/10.1016/j.phytochem.2007.07.006>
- Davies KM (1993) A cDNA clone for flavanone 3-hydroxylase from *Malus*. *Plant Physiol* 103(1):291
- Davies KM, Schwinn KE, Deroles SC, Manson DG, Lewis DH, Bloor SJ, Bradley JM (2003) Enhancing anthocyanin production by altering competition for substrate between flavonol synthase and dihydroflavonol 4-reductase. *Euphytica* 131:259–268
- Deboo GB, Albertsen MC, Taylor LP (1995) Flavanone 3-hydroxylase transcripts and flavonol accumulation are temporally coordinate in maize anthers. *Plant J* 7(5):703–713
- Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD, Bombarely A, Allen F, Hurst R, White B, Kernodle SP, Bromley JR, Sanchez-Tamburrino JP, Lewis RS, Mueller LA (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics* 18(1):448. <https://doi.org/10.1186/s12864-017-3791-6>
- Falcone Ferreyra ML, Rius S, Emiliani J, Pourcel L, Feller A, Morohashi K, Casati P, Grotewold E (2010) Cloning and characterization of a UV-B-inducible maize flavonol synthase. *Plant J* 62(1):77–91. <https://doi.org/10.1111/j.1365-313X.2010.04133.x>
- Falcone Ferreyra ML, Casas MI, Questa JI, Herrera AL, Deblasio S, Wang J, Jackson D, Grotewold E, Casati P (2012) Evolution and expression of tandem duplicated maize flavonol synthase genes. *Front Plant Sci* 3:101. <https://doi.org/10.3389/fpls.2012.00101>
- Falnes PO, Johansen RF, Seeberg E (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* 419(6903):178–182. <https://doi.org/10.1038/nature01048>
- Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztanyi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SC, Wu CH, Xenarios I, Yeh LS, Young SY, Mitchell AL (2017) InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res* 45(D1):D190–D199. <https://doi.org/10.1093/nar/gkw1107>
- Forkmann G, Heller W, Grisebach H (1980) Anthocyanin biosynthesis in flowers of *Matthiola incana* flavanone 3- and flavonoid 3'-hydroxylases. *Z Naturforsch Sect C Biosci* 35:691–695
- Fujita A, Goto-Yamamoto N, Aramaki I, Hashizume K (2006) Organ-specific transcription of putative flavonol synthase genes of grapevine and effects of plant hormones and shading on flavonol biosynthesis in grape berry skins. *Biosci Biotechnol Biochem* 70(3):632–638
- Geleijnse JM, Hollman P (2008) Flavonoids and cardiovascular health: which compounds, what mechanisms? *Am J Clin Nutr* 88(1):12–13
- Gronquist M, Bezzerides A, Attygalle A, Meinwald J, Eisner M, Eisner T (2001) Attractive and defensive functions of the ultraviolet pigments of a flower (*Hypericum calycinum*). *Proc Natl Acad Sci USA* 98(24):13745–13750. <https://doi.org/10.1073/pnas.231471698>
- Gueux N, Peitsch MC, Schwede T (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* 30(Suppl 1):S162–173. <https://doi.org/10.1002/elps.200900140>
- Harborne JB, Williams CA (2000) Advances in flavonoid research since 1992. *Phytochemistry* 55(6):481–504
- Himi E, Maekawa M, Noda K (2011) Differential expression of three flavanone 3-hydroxylase genes in grains and coleoptiles of wheat. *Int J Plant Genomics* 2011:369460. <https://doi.org/10.1155/2011/369460>
- Holton TA, Brugliera F, Tanaka Y (1993) Cloning and expression of flavonol synthase from *Petunia hybrida*. *Plant J* 4(6):1003–1010
- Ivan M, Kondo K, Yang H, Kim W, Valiano J, Ohh M, Salic A, Asara JM, Lane WS, Kaelin WG Jr (2001) HIF $\alpha$  targeted for VHL-mediated destruction by proline hydroxylation: implications for O<sub>2</sub> sensing. *Science* 292(5516):464–468. <https://doi.org/10.1126/science.1059817>
- Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23(10):403–405
- Kale A, Gawande S, Kotwal S (2008) Cancer phytotherapeutics: role for flavonoids at the cellular level. *Phytother Res* 22(5):567–577. <https://doi.org/10.1002/ptr.2283>
- Kang HK, Ecklund D, Liu M, Datta SK (2009) Apigenin, a non-mutagenic dietary flavonoid, suppresses lupus by inhibiting autoantigen presentation for expansion of autoreactive Th1 and Th17 cells. *Arthritis Res Ther* 11(2):R59. <https://doi.org/10.1186/ar2682>



- Kaur P, Shukla S, Gupta S (2008) Plant flavonoid apigenin inactivates Akt to trigger apoptosis in human prostate cancer: an in vitro and in vivo study. *Carcinogenesis* 29(11):2210–2217. <https://doi.org/10.1093/carcin/bgn201>
- Kawai Y, Ono E, Mizutani M (2014) Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J* 78(2):328–343. <https://doi.org/10.1111/tjp.12479>
- Kuhn BM, Geisler M, Bigler L, Ringli C (2011) Flavonols accumulate asymmetrically and affect auxin transport in *Arabidopsis*. *Plant Physiol* 156(2):585–595. <https://doi.org/10.1104/pp.111.175976>
- Kusano M, Tohge T, Fukushima A, Kobayashi M, Hayashi N, Otsuki H, Kondou Y, Goto H, Kawashima M, Matsuda F, Niida R, Matsui M, Saito K, Fernie AR (2011) Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of *Arabidopsis* to UV-B light. *Plant J* 67(2):354–369. <https://doi.org/10.1111/j.1365-313X.2011.04599.x>
- Lee HJ, Lloyd MD, Harlos K, Clifton IJ, Baldwin JE, Schofield CJ (2001) Kinetic and crystallographic studies on deacetoxycephalosporin C synthase (DAOCS). *J Mol Biol* 308(5):937–948. <https://doi.org/10.1006/jmbi.2001.4649>
- Lempe J, Balasubramanian S, Sureshkumar S, Singh A, Schmid M, Weigel D (2005) Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLoS Genet* 1(1):109–118. <https://doi.org/10.1371/journal.pgen.0010006>
- Lewis DR, Ramirez MV, Miller ND, Vallabhaneni P, Ray WK, Helm RF, Winkel BS, Muday GK (2011) Auxin and ethylene induce flavonol accumulation through distinct transcriptional networks. *Plant Physiol* 156(1):144–164. <https://doi.org/10.1104/pp.111.172502>
- Lin GZ, Lian YJ, Ryu JH, Sung MK, Park JS, Park HJ, Park BK, Shin JS, Lee MS, Cheon CI (2007) Expression and purification of histagged flavonol synthase of *Camellia sinensis* from *Escherichia coli*. *Protein Expr Purif* 55(2):287–292. <https://doi.org/10.1016/j.pep.2007.05.013>
- Maggioni D, Biffi L, Nicolini G, Garavello W (2014) Flavonoids in oral cancer prevention and therapy. *Eur J Cancer Prev* 24(6):517–528. <https://doi.org/10.1097/CEJ.0000000000000109>
- Marais G, Nouvellet P, Keightley PD, Charlesworth B (2005) Intron size and exon evolution in *Drosophila*. *Genetics* 170(1):481–485. <https://doi.org/10.1534/genetics.104.037333>
- Meldgaard M (1992) Expression of chalcone synthase, dihydroflavonol reductase, and flavanone-3-hydroxylase in mutants of barley deficient in anthocyanin and proanthocyanidin biosynthesis. *Theor Appl Genet* 83(6–7):695–706. <https://doi.org/10.1007/BF00226687>
- Menssen A, Hohmann S, Martin W, Schnable PS, Peterson PA, Saedler H, Gierl A (1990) The *En/Spm* transposable element of *Zea mays* contains splice sites at the termini generating a novel intron from a *dSpm* element in the *A2* gene. *EMBO J* 9(10):3051–3057
- Mo Y, Nagel C, Taylor LP (1992) Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. *Proc Natl Acad Sci USA* 89(15):7213–7217
- Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci USA* 107(10):4623–4628. <https://doi.org/10.1073/pnas.0907801107>
- Owens DK, Alerding AB, Crosby KC, Bandara AB, Westwood JH, Winkel BS (2008) Functional analysis of a predicted flavonol synthase gene family in *Arabidopsis*. *Plant Physiol* 147(3):1046–1061. <https://doi.org/10.1104/pp.108.117457>
- Pelletier MK, Shirley BW (1996) Analysis of flavanone 3-hydroxylase in *Arabidopsis* seedlings. Coordinate regulation with chalcone synthase and chalcone isomerase. *Plant Physiol* 111(1):339–345
- Prescott AG, Lloyd MD (2000) The iron (II) and 2-oxoacid-dependent dioxygenases and their role in metabolism. *Nat Prod Rep* 17(4):367–383
- Prescott AG, Stamford NP, Wheeler G, Firmin JL (2002) In vitro properties of a recombinant flavonol synthase from *Arabidopsis thaliana*. *Phytochemistry* 60(6):589–593
- Preuss A, Stracke R, Weisshaar B, Hillebrecht A, Matern U, Martens S (2009) *Arabidopsis thaliana* expresses a second functional flavonol synthase. *FEBS Lett* 583(12):1981–1986. <https://doi.org/10.1016/j.febslet.2009.05.006>
- Routaboul JM, Dubos C, Beck G, Marquis C, Bidzinski P, Loudet O, Lepiniec L (2012) Metabolite profiling and quantitative genetics of natural variation for flavonoids in *Arabidopsis*. *J Exp Bot* 63(10):3749–3764. <https://doi.org/10.1093/jxb/ers067>
- Saito K, Kobayashi M, Gong Z, Tanaka Y, Yamazaki M (1999) Direct evidence for anthocyanidin synthase as a 2-oxoglutarate-dependent oxygenase: molecular cloning and functional expression of cDNA from a red form of *Perilla frutescens*. *Plant J* 17(2):181–189
- Sierro N, van Oeveren J, van Eijk MJ, Martin F, Stormo KE, Peitsch MC, Ivanov NV (2013) Whole genome profiling physical map and ancestral annotation of tobacco Hicks Broadleaf. *Plant J* 75(5):880–889. <https://doi.org/10.1111/tjp.12247>
- Sierro N, Battey JN, Ouadi S, Bakaher N, Bovet L, Willig A, Goepfert S, Peitsch MC, Ivanov NV (2014) The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun* 5:3833. <https://doi.org/10.1038/ncomms4833>
- Stracke R, De Vos RC, Bartelniewoehner L, Ishihara H, Sagasser M, Martens S, Weisshaar B (2009) Metabolomic and genetic analyses of flavonol synthesis in *Arabidopsis thaliana* support the in vivo involvement of leucoanthocyanidin dioxygenase. *Planta* 229(2):427–445. <https://doi.org/10.1007/s00425-008-0841-y>
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–2739. <https://doi.org/10.1093/molbev/msr121>
- Treutter D (2005) Significance of flavonoids in plant resistance and enhancement of their biosynthesis. *Plant Biol (Stuttg)* 7(6):581–591. <https://doi.org/10.1055/s-2005-873009>
- Trewick SC, Henshaw TF, Hausinger RP, Lindahl T, Sedgwick B (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* 419(6903):174–178. <https://doi.org/10.1038/nature00908>
- Turnbull JJ, Nakajima J, Welford RW, Yamazaki M, Saito K, Schofield CJ (2004) Mechanistic studies on three 2-oxoglutarate-dependent oxygenases of flavonoid biosynthesis: anthocyanidin synthase, flavonol synthase, and flavanone 3beta-hydroxylase. *J Biol Chem* 279(2):1206–1216. <https://doi.org/10.1074/jbc.M309228200>
- van Dam RM, Naidoo N, Landberg R (2013) Dietary flavonoids and the development of type 2 diabetes and cardiovascular diseases: review of recent findings. *Curr Opin Lipidol* 24(1):25–33. <https://doi.org/10.1097/MOL.0b013e32835bcdff>
- Welford RW, Turnbull JJ, Claridge TD, Prescott AG, Schofield CJ (2001) Evidence for oxidation at C-3 of the flavonoid C-ring during anthocyanin biosynthesis. *Chem Commun (Camb)* 18:1828–1829
- Welford RW, Clifton IJ, Turnbull JJ, Wilson SC, Schofield CJ (2005) Structural and mechanistic studies on anthocyanidin synthase catalysed oxidation of flavanone substrates: the effect of C-2 stereochemistry on product selectivity and mechanism. *Org Biomol Chem* 3(17):3117–3126. <https://doi.org/10.1039/b507153d>
- Wellmann F, Matern U, Lukacin R (2004) Significance of C-terminal sequence elements for *Petunia flavanone* 3β-hydroxylase activity. *FEBS Lett* 561(1–3):149–154. [https://doi.org/10.1016/S0014-5793\(04\)00159-0](https://doi.org/10.1016/S0014-5793(04)00159-0)
- Wellmann F, Griesser M, Schwab W, Martens S, Eisenreich W, Matern U, Lukacin R (2006) Anthocyanidin synthase from *Gerbera hybrida* catalyzes the conversion of (+)-catechin to cyanidin and



- a novel procyanidin. FEBS Lett 580(6):1642–1648. <https://doi.org/10.1016/j.febslet.2006.02.004>
- Williams CA, Grayer RJ (2004) Anthocyanins and other flavonoids. Nat Prod Rep 21(4):539–573. <https://doi.org/10.1039/b311404j>
- Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol 126(2):485–493
- Xu F, Li L, Zhang W, Cheng H, Sun N, Cheng S, Wang Y (2012) Isolation, characterization, and function analysis of a flavonol synthase gene from *Ginkgo biloba*. Mol Biol Rep 39(3):2285–2296. <https://doi.org/10.1007/s11033-011-0978-9>