DR YUKI TOBIMATSU (Orcid ID : 0000-0002-7578-7392)

DR CLIVE LO (Orcid ID : 0000-0002-1995-3843)

Article type : MS - Regular Manuscript

Full title: Convergent recruitment of 5′-hydroxylase activities by CYP75B flavonoid B-ring hydroxylases for tricin biosynthesis in *Medicago* legumes

Authors/affiliations: Andy C W Lui[1], Pui Ying Lam[2], Kwun Ho Chan[1], Lanxiang Wang[1], Yuki Tobimatsu[2], Clive Lo[1]

[1]School of Biological Sciences, The University of Hong Kong, Hong Kong, China

[2]Research Institute for Sustainable Humanosphere, Kyoto University, Uji, Kyoto 611-0011, Japan

ORCID: ACWL (0000-0002-9604-7676); PYL (0000-0001-5025-1308); LW (0000-0002-4872-9824); YT (0000-0002-7578-7392); CL (0000-0002-1995-3843)

Corresponding author: Clive Lo (clivelo@hku.hk; +852-2299-0337)

Received: 5 August 2019

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/NPH.16498

Accepted: 14 February 2020

Word count

Main text total word count: 6499

Introduction: 981

Materials and methods: 1366

Results: 2269

Discussion: 1883

Acknowledgements: 87

Author contributions: 31

Number of figures: 5 (Fig. 4 & 5 should be published in color)

Number of tables: 2

Number of supporting materials: 12

Summary

- Tricin (3′,5′-dimethoxylated flavone) is a predominant flavonoid amongst monocots but occurs only in isolated and unrelated dicot lineages. Although tricin biosynthesis has been intensively studied in monocots, it remained largely elusive in tricin-accumulating dicots.
- We investigated a subgroup of cytochrome P450 (CYP) 75B subfamily flavonoid B-ring hydroxylases (FBHs) from two tricin-accumulating legumes, *Medicago truncatula* and alfalfa (*M. sativa*), by phylogenetic, molecular, biochemical and mutant analyses.
- Five *Medicago* cytochrome P450 CYP75B FBHs are phylogenetically distant from other legume

CYP75B members. Among them, *MtFBH-4*, *MsFBH-4* and *MsFBH-10* were expressed in tricin-accumulating vegetative tissues. *In vitro* and *in planta* analyses demonstrated that these proteins catalyze 3′- and 5′-hydroxylations critical to tricin biosynthesis. A key amino acid polymorphism, T492G, at their Substrate Recognition Site 6 domain is required for the novel 5′-hydroxylation activities. *M. truncatula mtfbh-4* mutants were tricin-deficient, indicating that MtFBH-4 is indispensable for tricin biosynthesis.

- Our results revealed that these *Medicago* legumes had acquired the tricin pathway through molecular evolution of CYP75B FBHs subsequent to speciation from other non-tricin-accumulating legumes. Moreover, their evolution is independent from that of grass-specific CYP75B apigenin 3′-hydroxylases/chrysoeriol 5′-hydroxylases dedicated to tricin production and Asteraceae CYP75B flavonoid 3′,5′-hydroxylases catalyzing the production of delphinidin-based pigments.

Keywords: Alfalfa (*Medicago sativa*), dicot, legume, *Medicago truncatula*, pathway evolution, tricin biosynthesis.

Introduction

Flavonoids are a group of secondary metabolites widespread among land plants. These structurally diverse compounds contribute to various physiological mechanisms in plants, including responses towards biotic and abiotic threats, nodule organogenesis (in legumes), attraction of pollinators, fertility, pigmentation of fruits and flowers and regulation of plant developments (Dixon & Pasinetti, 2010; Falcone Ferreyra *et al*., 2012; Harborne & Williams, 2000; Peer & Murphy, 2006; Zhang *et al*., 2009; Zhang *et al*., 2007). In addition, some flavonoids are well-documented nutraceuticals. For instance, tricin, a 3′,5′-dimethoxylated flavone, demonstrates potent anticancer (Al-Fayez *et al*., 2006; Cai *et al*., 2004), antiproliferative (Duarte-Almeida *et al*., 2007) and

antioxidant (Kwon *et al*., 2002) properties. More recent studies showed that tricin is an integrated component of lignin, which is an abundant phenylpropanoid polymer in cell walls, in some plant lineages (Del Río *et al*., 2012; Lan *et al*., 2015; Lan *et al*., 2016). Hence, there is a tight but complex association between flavonoid and lignin biosynthesis, prompting efforts in modifying straw lignin content and composition through manipulation of tricin biosynthesis (Eloy *et al*., 2017; Lam *et al*., 2017; Lam *et al*., 2019a; Lam *et al*., 2019b). Consequently, a thorough understanding of tricin biosynthesis in species with great potential in biomass utilization, such as sorghum (*Sorghum bicolor*) and alfalfa (*Medicago sativa*) (Lan *et al*., 2016), has become imperative.

In monocots, tricin derivatives including soluble tricin *O*-conjugates and cell-wall-integrated tricin-lignins are ubiquitously present in major families like Poaceae, Arecaceae, Cyperaceae, and Orchidaceae (Lan *et al*., 2016; Li *et al*., 2016). Considering the more extensive occurrence of tricin metabolites in monocots, prior investigation of the enzymology of its biosynthesis focused primarily on Poaceae, including cereal crops like rice, maize and sorghum (Eloy *et al*., 2017; Lam *et al*., 2015; Lam *et al*., 2019a; Lam *et al*., 2019b). Chalcone synthase, the first committed enzyme of flavonoid biosynthesis, condensates three malonyl-CoA with one *p*-coumaroyl-CoA into naringenin chalcone which is swiftly converted into naringenin either catalyzed by chalcone isomerase or via spontaneous isomerization (Druka *et al*., 2003; Mol *et al*., 1985; Reddy *et al*., 1996). Flavone synthase I and/or flavone synthase II (FNSII) then desaturate the C2–C3 bond of naringenin to generate apigenin (Falcone Ferreyra *et al*., 2015; Lam *et al*., 2017; Lam *et al*., 2014; Martens & Mithofer, 2005; Zhang *et al*., 2007). Afterwards, the apigenin aglycone is subjected to two rounds of sequential hydroxylation and methylation at the 3′- and 5′- positions in the flavonoid B-ring to complete the biosynthesis of tricin in grasses (Cummins *et al*., 2006; Kim *et al*., 2006; Zhou *et al*., 2006; Zhou *et al*., 2009). Interestingly, in contrast to the earlier belief that a canonical flavonoid 3′,5′-hydroxylase (F3′5′H) belonging to CYP75A subfamily is required for the 5′-hydroxylation step (Zhou & Ibrahim, 2010), the only member in rice (CYP75A11) appeared to be non-functional (Lam *et al*., 2015). Instead, apigenin 3′-hydroxylase/chrysoeriol 5′-hydroxylase (A3′H/C5′H) from CYP75B subfamily (to which canonical F3′Hs belong) serves as a bifunctional flavone hydroxylase which catalyzes both the 3′- and 5′-hydroxylation steps exclusively for the biosynthesis of soluble and lignin-integrated tricin in grasses (Lam *et al*., 2015; Lam *et al*., 2019a).

In contrast to the widespread occurrence in monocots, the distribution of tricin-type metabolites is more sporadic in dicots. They have been detected in isolated species of unrelated lineages scattering across dicot families such as Ranunculaceae (peripheral dicots), Fabaceae (rosids) and Caprifoliaceae (asterids) (Li *et al*., 2016). Unlike Poaceae, tricin biosynthesis has remained largely uncharacterized in dicots, but the patchy occurrences may suggest that independent pathways evolved convergently in different tricin-accumulating lineages. *Medicago* species (e.g., *M. sativa* or alfalfa and *M. truncaluta*) are rare examples of tricin-accumulating plants in Fabaceae (Li *et al*., 2016; Zhou & Ibrahim, 2010). Metabolomic studies have provided comprehensive data for the accumulation of tricin and its upstream flavone intermediates (i.e., apigenin, luteolin and chrysoeriol) in these *Medicago* legumes (Fu & Wang, 2015; Jasinski *et al*., 2009; Kowalska *et al*., 2007). It has been demonstrated that in *M. truncatula*, apigenin is produced from naringenin via the formation of 2-hydroxynaringenin as an intermediate instead of directly introducing a C2=C3 bond into the flavanone skeleton as in Poaceae (Supporting Information Fig. **S1**) (Akashi *et al*., 1999; Du *et al*., 2010; Zhang *et al*., 2007). However, downstream metabolic steps and enzymes that further derivatize apigenin into tricin, especially those related to 5′-hydroxylation, are still ambiguous. While CYP75B A3′H/C5′Hs were identified as the *bona fide* enzymes catalyzing this reaction step in grasses, these flavonoid B-ring hydroxylases (FBHs) are strictly conserved in Poaceae (Lam *et al*., 2015; Lam *et al*., 2019a). Hence, *Medicago* spp. must have independently recruited other forms of FBHs for tricin biosynthesis. In particular, it remained unclear whether the 5′-hydroxylation involves canonical CYP75A F3′5′Hs (Dixon & Steele, 1999; Marles *et al*., 2003; Moreau *et al*., 2012), or CYP75B FBHs with additional 5′-hydroxylase activities similar to the A3′H/C5′Hs in grasses (Lam *et al*., 2015; Lam *et al*., 2019a).

In this study, we analyzed *M. truncatula* and alfalfa FBHs and identified a unique subgroup of CYP75B enzymes that are responsible for catalyzing both 3′- and 5′-hydroxylations essential to tricin biosynthesis under *in vitro* and *in planta* conditions. The recruitment of flavonoid 5′-hydroxylation activities of these *Medicago* CYP75B FBHs is strongly correlated with an amino acid polymorphism T492G in the substrate recognition site 6 (SRS6) motif. Furthermore, two independent *M. truncatula* *MtFBH-4* mutant lines displayed a tricin-deficient phenotype in their vegetative tissues. Altogether our findings strongly support that unique FBHs were recruited for tricin biosynthesis in the genus

Medicago, and such information provides valuable insights into future manipulation of tricin-metabolites, including cell-wall-bound tricin-lignin in alfalfa for improvement of forage quality. The evolution and acquisition of the 5′-hydroxylase activities by this subgroup of *Medicago*-specific CYP75B FBHs are independent from the Poaceae-family-specific A3′H/C5′Hs, which are also involved in tricin biosynthesis. In addition, genus-specific convergent evolutionary events plausibly contribute to the patchy occurrence of tricin-type metabolites in dicots today.

Materials and methods

Phylogenetic analysis of FBHs in the *Medicago* genus

To identify putative FBHs in the *Medicago* genus, the protein sequence of a previously described soybean F3′H (*Glycine max*; BAB83261.1; Toda *et al*., 2002) was BLASTP searched against the protein sequence databases of *M. truncatula* (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/) and alfalfa (Noble Research Institute, LLC, http://www.alfalfatoolbox.org/). The retrieved sequences were aligned with known FBHs from both dicots and monocots using ClustalW with default configurations (Larkin *et al*., 2007). An unrooted phylogenetic tree was constructed by the neighbor-joining method (JTT model) with 1000 bootstrap replicates using MEGA7 (Jones *et al*., 1992; Kumar *et al*., 2016).

Plant materials and *M. truncatula tnt1* insertion mutants

Seeds of two independent *M. truncatula* mutant lines for *MtFBH-4* [NF17189 (R1; *mtfbh-4-a*) and NF19367 (R0; *mtfbh-4-b*)] which harbor transposable element of *Nicotiana tabacum* cell type 1

(*tnt1*) insertions were obtained from the *M. truncatula tnt1* insertion mutant library (Tadege *et al*., 2008; available at Noble Research Institute, LLC, http://medicago-mutant.noble.org/mutant/). Seeds of *M. truncatula* wild type (cv. R108-1), *tnt1* mutants and alfalfa wild type were germinated on 0.8% (w/v) water agar. Germinated seeds were then sown into soil and incubated in a phytotron (12/12 h light/dark cycle, 24 °C). Homozygous *M. truncatula MtFBH-4 tnt1* insertion mutants were confirmed by genomic PCR and RT-PCR using *tnt1*-specific primers and *MtFBH4*-specific primers (Supporting Information Table **S3**) as described previously (Cheng *et al*., 2011).

Gene expression analyses experiments

In silico expression data of *MtFBH* genes were retrieved from the MedicMine database (Krishnakumar *et al*., 2015; J. Craig Venter Institute, http://medicmine.jcvi.org/medicmine/). Total RNA was extracted from leaves and stems of 1-month-old alfalfa and 2-month-old *M. truncatula* cv. R108-1 and cv. A17 plants using the TRIzol method (Invitrogen) according to the manufacturer´s instructions. Total RNA (1µg) was reverse transcribed into first-strand cDNA using an oligo-dT primer and the PrimeScript RT reagent Kit with gDNA Eraser (TaKaRa). *In planta* expression of different genes in *M. truncatula* cv. R108 (n = 3) was then examined by qRT-PCR using ABI StepOne Plus Real-Time PCR system (Applied Biosystems) and gene-specific primers (Supporting Information Table **S3**). A housekeeping gene *Medicago truncatula polyubiquitin 3* (*MtUbi*; XM_003627103.1; Kakar *et al*., 2008) was used as an internal control.

Cloning of *Medicago* CYP75B *FBH* genes

The full length coding sequence (CDS) of *MtFBH-4* was amplified from cDNA prepared from the A17 ecotype (Supporting Information Table **S2**). Full-length CDS of *MtFBH-1/2/5* and *MsFBH-4/10* were chemically synthesized (Synbio Technologies), followed by amplification using gene-specific primers (Supporting Information Table **S2**) and ligation into the pYES2.1/V5-His-TOPO yeast expression vector (Invitrogen).

Generation of MtFBH-4 mutant protein

The full-length CDS of *MtFBH-4* from the A17 ecotype was chemically synthesized (Synbio

Technologies) to introduce a specific point mutation at the amino acid sequence level. The nucleotide triplet, GGT, from the 1474th to 1476th position of the *MtFBH-4* CDS was replaced by another set of nucleotide triplet, ACT, in the synthesized sequence. The PCR products amplified from the synthesized DNA template were cloned into the pYES2.1/V5-His-TOPO expression vector (Invitrogen). Consequently, the original Gly residue at the 492nd position was replaced by a Thr residue, resulting in a nonsynonymous mutation (G492T) in the mutated protein sequence.

Heterologous expression in yeast and *in vitro* enzyme assays

Yeast expression constructs were transformed into a yeast (*Saccharomyces cerevisiae*) strain INVSc1 (Invitrogen) using the LiAc/PEG method (Gietz and Schiestl, 2007). Yeast incubation, induction of gene expression and yeast microsome extraction were carried out as described by Du *et al*. (2010). Gene expression in the transformed yeast cells was verified by RT-PCR. Microsomes expressing recombinant CYP75A1 protein from *Petunia hybrida* (Lam *et al*., 2015) were used as a positive control for all enzyme reactions. To investigate the enzyme activities of the *M. truncatula* and alfalfa FBHs, individual microsome preparations harboring 200 μg proteins were incubated in 100 mM potassium phosphate buffer (pH 7.0), 5 mM NADPH, 5 mM L-glutathione and 100 μM flavonoid substrates at 30 °C for 1 h. Enzyme kinetics of recombinant MtFBH-4 and MtFBH-5 were determined under the same conditions except the concentration of apigenin used ranged from 1 to 100 μM. Initial velocities were plotted against different apigenin concentrations using the non-linear regression fitting from the GraphPad Prism software (GraphPad). All reactions were repeated three times using the same microsome preparation. To compare the 3′- and 5′-hydroxylation activities between the wild-type and mutant MtFBH-4 proteins, four different flavonoids (apigenin, chrysoeriol, homoeriodictyol and isorhamnetin) were incubated with microsome preparations harboring the two proteins individually. Four sets of microsomes (2 containing the wild-type protein from the A17 ecotype and 2 containing the point mutation protein, each from an independent yeast colony), were individually incubated with the flavonoid substrates under the standard assay conditions described above and all reactions were repeated 3 times. Each enzyme assay reaction mixture was extracted twice by ethyl acetate and the combined organic layer was dried and re-suspended in 50 μl HPLC-grade methanol for HPLC-quadrupole time-of-flight mass spectrometer (QTOF-MS) analysis as

described below.

Agroinfiltration of *N. benthamiana* leaves

The full-length CDS of *MtFBH-1/2/4/5* and *MsFBH-4*/*10* were subcloned into a binary vector, pEAQ-HT (Sainsbury *et al*., 2009), which were then transformed into *Agrobacterium tumefaciens* strain GV3101 by the freeze-thawed method (Holsters *et al*., 1978). The obtained *Agrobacterium* suspensions were pressure-infiltrated into the abaxial side of 4-week-old *N. benthamiana* leaves as previously described (Sainsbury *et al*., 2009). *Agrobacterium* harboring the empty pEAQ-HT vector was simultaneously infiltrated into *N. benthamiana* leaves as experimental controls. Four days after the infiltration, *N. benthamiana* leaves were harvested and immediately vacuumed in an aqueous solution (with 3% (v/v) dimethyl sulfoxide to enhance substrate solubility) of 100 μM apigenin or chrysoeriol for 30 min and were then incubated at room temperature for 24 h. After incubation, the infiltrated *N. benthamiana* leaves were washed twice with deionized water, dried on filter papers, and their metabolites were extracted by methanol for HPLC-tandem mass spectrometer (MS/MS) analysis after acid-hydrolysis as described below.

Plant metabolite extraction

N. benthamiana leaves (200 mg) transiently expressing *MtFBH-1*, *MtFBH-2*, *MtFBH-4*, *MtFBH-5*, *MsFBH-4* or *MsFBH-10*, as well as stems and leaves (100 mg) of 2-month old *M. truncatula* wild type (cv. R108-1) and *MtFBH-4 tnt1* mutants (n = 5 for each line) were harvested for metabolite analysis. Tissues were frozen by liquid nitrogen, ground into fine powder using TissueLyser II (QIAGEN) and were then extracted with 200 μl HPLC-grade methanol at 4 °C for 24 h. For acid hydrolysis, filtered samples were incubated with an equal volume of 2 M HCl at 90 °C for 1 h.

HPLC-QTOF-MS and HPLC-MS/MS analysis

Purified products of enzyme assays were separated by a Synergi C18 column (Synergi 4μ Fusion RP 80 A, 50 × 2 mm, Phenomenex) connected to a QTOF-mass spectrometer X500R system (AB SCIEX) operating under positive ionization mode. The mobile phase consisted of 0.5% (v/v)

formic acid/water (A) and 0.5% (v/v) formic acid/methanol (B). The flow rate was maintained at 0.5 ml min^{-1} over a 6-min linear gradient of 10% to 70% B. Compound detection was achieved by information dependent acquisition (IDA) mode. Meanwhile, acid-hydrolyzed plant metabolites were separated by the same C18 column connected to a HP1100 series HPLC system (Agilent Technologies) linked to an AP3200-QTRAP mass spectrometer (AB SCIEX) operating under positive ionization mode. A linear gradient from 10% to 70% B over 20 min at an 0.2 ml min^{-1} flow rate was used. MS/MS spectra were acquired by Enhanced Product Ion scan. Total non-acid-hydrolyzed metabolites were analyzed by the X500R QTOF-MS system operating under positive ionization mode after compound separation on a Kinetex F5 column (Kinetex 2.6 μm F5 100 A, 150 × 2.1 mm, Phenomenex) over a linear gradient of 5% to 95% B in 20 min at a flow rate of 0.2 ml min^{-1}. Quantification of metabolites was performed by integration of peak area using the quantification mode of SCIEX OS. Identification of compounds was achieved via comparing the retention time and MS/MS spectra with previously published data or authentic standards (Du *et al*., 2010; Fu & Wang, 2015; Kowalska *et al*., 2007; Lam *et al*., 2015).

Accession numbers

Sequence data from this article can be found in the GenBank data library, Phytozome or the alfalfa database under accession numbers as listed in Supporting Information Table **S1** and **S2**.

Results

In silico analysis identifies a unique group of FBHs in *Medicago* spp.

Seven putative MtFBHs (two CYP75A and five CYP75B sequences) in *M. truncatula* and fourteen putative MsFBHs (three CYP75A and eleven CYP75B sequences) in alfalfa were retrieved (Supporting Information Table **S1**) following BLASTP searches against the respective proteomes using a known soybean GmF3′H sequence (Toda *et al*., 2002). Since MtFBH-3 is truncated by approximately 240 amino acids at the N-terminus compared to other FBHs, it was considered an incomplete protein and removed from subsequent analyses.

Multiple sequence alignment of the *Medicago* FBHs revealed the conservation of classical cytochrome P450 domains including an ExxR motif, a heme binding domain and an oxygen binding pocket (data not shown). In addition, most of the SRSs are highly conserved except SRS6, a key region determining substrate specificity of cytochrome P450 enzymes (Dueholm *et al*., 2015; Gotoh, 1992). Interestingly, at the 8th position of SRS6 (SRS6-8), a Gly residue is present in MtFBH-4, MsFBH-1, MsFBH-3, MsFBH-4, and MsFBH-10 (Table **1**) instead of a Thr residue which is highly conserved in canonical CYP75B F3′Hs (Han *et al*., 2010; Schoenbohm *et al*., 2000; Toda *et al*., 2002), or an Ala residue which is highly conserved in canonical CYP75A F3′5′Hs (Seitz *et al*., 2007).

Phylogenetic analysis revealed that the five FBHs with a Gly residue at SRS6-8 formed a separate clade from the other *Medicago* CYP75B FBHs with a Thr residue (Table **1** & Fig. **1**). Recently, we reported a unique subgroup of grass-specific CYP75B proteins with A3´H/C5´H activities required for 3´- and 5´-hydroxylations during tricin biosynthesis in grasses (Lam *et al*., 2019a). Coincidentally, a Leu residue, rather than a Thr/Ala residue, is present at SRS6-8 in all grass A3´H/C5´Hs characterized, and they also formed a distinct clade from the canonical F3´Hs in grasses (Fig. **1**). Therefore, we hypothesized that the *Medicago* CYP75B FBHs with a Gly residue at SRS6-8 (MtFBH-4, MsFBH-1, MsFBH-3, MsFBH-4, and MsFBH-10) may display unique catalytic activities required for tricin biosynthesis, similar to the grass CYP75B A3´H/C5´Hs.

Gene expression analyses implicate an involvement of MtFBH-4 in tricin biosynthesis

Our investigation on the putative *Medicago* FBHs was initiated with *M. truncatula*, which is a well-recognized diploid model legume closely related to commercially important alfalfa (an autotetraploid with high genetic complexity) with a short life cycle, small genome and publicly available genomic and genetic resources (Tadege *et al*., 2008). We first examined the expression patterns of the putative *MtFBH*s in *M. truncatula* leaves and stems which were reported to accumulate tricin-related metabolites (Fu & Wang, 2015; Li *et al*., 2016). Through analyzing publicly available *in silico* expression datasets (Krishnakumar *et al*., 2015) and confirming these data by qRT-PCR experiments (Supporting Information Fig. **S2a-b**), we determined that genes encoding the two CYP75A members, i.e., *MtFBH-6*/*7* (Supporting Information Table **S1**), were either poorly expressed (*MtFBH-6*) or not expressed (*MtFBH-7*) in *M. truncatula* leaves and stems, suggesting that these putative canonical F3´5´Hs are not involved in tricin biosynthesis in vegetative tissues of *M. truncatula*. Among the CYP75B members, i.e., *MtFBH-1/2/4/5* (Supporting Information Table **S1**), only *MtFBH-4*/*5* were consistently expressed in both leaves and stems; the expression levels of *MtFBH-1* and *MtFBH-2* were very low or barely detectable in *M. truncatula* leaves and stems (Supporting Information Fig. **S2a-b**). Considering the expression pattern, unique SRS6-8 amino acid residue (Table **1**) and phylogenetic group of *MtFBH-4* (Fig. **1**), we reasoned that MtFBH-4 might be responsible for catalyzing the 5´-hydrxoylation step during tricin biosynthesis in *M. truncatula*.

***In vitro* and *in planta* enzyme assays reveal unique activities of MtFBH-4**

Subsequently, *MtFBH-4* was heterologously expressed in yeast cells and microsome fractions harboring the recombinant proteins were purified and incubated with different flavonoid substrates for enzymatic assays. HPLC-QTOF-MS analysis of the reaction products demonstrated that the recombinant protein catalyzed 3′-hydroxylations of flavone (apigenin to luteolin, Fig. **2b**), flavonol (kaempferol to quercetin, Supporting Information Fig. **S3a**), and flavanone (naringenin to eriodictyol, Supporting Information Fig. **S3b**), but not 3′-hydroxylation of dihydroxyflavonol (dihydrokaempferol to dihydroxyquercetin, Supporting Information Fig. **S3c**). In contrast, microsomes expressing a previously characterized dicot F3′5′H, i.e., petunia CYP75A1 (Holton *et al*., 1993), were able to introduce a 3′-hydroxyl group to the all flavonoids tested including dihydrokaempferol (Supporting Information Fig. **S3a-c**).

More importantly, consistent with our hypothesis, recombinant MtFBH-4 was able to convert chrysoeriol into selgin via introducing a 5′-hydroxyl group in the flavone B-ring (Fig. **2c**), a key enzymatic activity distinguishing the grass A3′H/C5′Hs from other CYP75B F3′Hs in monocots (Lam *et al*., 2015). Moreover, its flavonoid 5′-hydroxylation activities were further extended to two other 3′-methoxylated flavonoid substrates, homoeriodictyol and isorhamnetin, since 5′-hydroxyhomoeriodictyol and laricitrin were detected respectively as their enzyme reaction products (Fig. **2d-e**). On the other hand, 3′-hydroxylated flavonoids such as luteolin and quercetin (Supporting Information Fig. **S3d-e**) were not modified by MtFBH-4 whereas petunia CYP75A1 F3′5′H was able to convert these substrates into 3′,5′-hydroxylated flavonoids under the same conditions. These data indicated that under *in vitro* conditions, MtFBH-4 can 3′-hydroxylate a range of flavonoid substrates, whereas its 5′-hydroxylase activities are limited to 3′-methoxylated flavonoids. Its catalytic activities are different from those observed in CYP75A1, a canonical dicot F3′5′H, which modifies 3′-hydroxylated as well as 3′-methoxylated flavonoids, and grass CYP75B A3′H/C5′Hs, which modify chrysoeriol but not the other 3′-methoxylated flavonoids (Lam *et al*., 2015).

In addition, we analyzed the catalytic activities of other CYP75B FBHs in *M. truncatula* including MtFBH-1 and MtFBH-2, which are phylogenetically closely related to MtFBH-4 (Fig. **1**), as well as MtFBH-5, which is expressed in tricin-accumulating vegetative tissues (Supporting Information Fig. **S2**). While recombinant MtFBH-1, MtFBH-2 and MtFBH-5 could convert apigenin

into luteolin, they all failed to modify 3´-substituted flavone substrates including luteolin and chrysoeriol at the 5´-position under *in vitro* conditions (Supporting Information Fig. **S4a-c**), suggesting that MtFBH-4 may be the sole enzyme that can utilize chrysoeriol as a substrate for tricin biosynthesis in the vegetative tissues of *M. truncatula*.

MtFBH-4 was also transiently expressed in *Nicotiana benthamiana* leaves under the control of a highly efficient Cowpea Mosaic Virus expression cassette (Sainsbury *et al*., 2009) to evaluate its *in planta* enzymatic functions. *Agrobacterium*-infiltrated leaves (n = 5) expressing *MtFBH-4* were fed with apigenin or chrysoeriol and the resultant metabolites were extracted and acid-hydrolyzed to remove *O*-conjugates for HPLC-tandem mass spectrometer (MS/MS) analysis. As shown in Fig. **3a** and Supporting Information Fig. **S5a**, apigenin and chrysoeriol substrates could be detected in all the leaves transformed with empty vectors or constructs expressing *MtFBH-4*, indicating successful uptake of individual substrates by the leave tissues. In leaves expressing *MtFBH-4* infiltrated with apigenin, all flavones along the tricin biosynthetic pathway, including luteolin, chrysoeriol, selgin and tricin, were detected at high intensities while only trace amount of luteolin and chrysoeriol were detected in control leaves expressing the empty vector (Fig. **3b-e**), likely due to the endogenous F3´H and *O*-methyltransferase activities in the *N. benthamiana* leaves. Furthermore, leaves expressing *MtFBH-4* infiltrated with chrysoeriol accumulated selgin and tricin while these signals were absent in the control leaves expressing the empty vector (Supporting Information Fig. **S5a-c**); the conversion of selgin into tricin could be likewise a result of endogenous *O*-methyltransferase activities in the *N. benthamiana* leaves. On the other hand, we also analyzed the catalytic activities of MtFBH-1, MtFBH-2 and MtFBH-5 which are CYP75B FBHs with a Thr residue at SRS6-8 (Table **1**). Consistent with the *in vitro* enzyme assay results, after incubation with apigenin, leaves transiently expressing either *MtFBH-1*, *MtFBH-2* (Supporting Information Fig. **S6a-e)** or *MtFBH-5* (Fig. **3b-e**) accumulated luteolin and chrysoeriol, but not selgin or tricin, as end products. Altogether, these results suggested that under *in planta* conditions, MtFBH-4 alone is sufficient to catalyze both 3´and 5´-hydroxylation steps to the complete the tricin pathway while MtFBH-1/2/5 function as canonical F3´Hs under *in planta* conditions (Fig. **3f** & Supporting Information Fig. **S6f**).

SRS6 mutation of MtFBH-4 results in changes of its 5´-hydroxylation activities

We speculated that the Gly492 residue at SRS6-8 could be a contributing factor for the recruitment of 5′-hydroxylation activities in MtFBH-4. Accordingly, we generated a recombinant MtFBH-4 G492T protein by replacing the Gly492 residue with a Thr492 residue (Fig. **2a**). Yeast microsomes carrying either the wild-type MtFBH-4 protein (WT) or the G492T point mutant protein (PM) were incubated with apigenin, chrysoeriol, homoeriodictyol, or isorhamnetin. Purified enzyme reaction products were analyzed using HPLC-QTOF-MS (Fig. **2b-e**) and their relative product ion abundances were recorded and compared (Fig. **2f**). Interestingly, the replacement of the Gly492 residue in MtFBH-4 led to a 94% reduction in its 5′-hydroxylation activities towards chrysoeriol, in addition to the complete loss of catalytic activities towards homoeriodictyol and isorhamnetin without hampering its 3′-hydroxylation activity towards apigenin (Fig. **2b-f**). Overall, these results strongly support our notions that the Gly residue at SRS6-8 in MtFBH-4 is required for the 5′-hydroxylation of 3′-methoxylated flavonoid substrates, and that the incorporation of this amino acid polymorphism is likely one of the many molecular evolution events that eventually led to the independent evolution of tricin pathway in *M. truncatula*.

Defective MtFBH-4 abolishes tricin biosynthetic pathway in *M. truncatula*

To further elucidate the role of MtFBH-4 in the tricin biosynthetic pathway of *M. truncatula*, two *tnt1*-retrotransposon-tagged mutant lines, NF17189 (*mtfbh-4-a*) and NF19367 (*mtfbh-4-b*) (Fig. **4a**), identified from the *Medicago truncatula tnt1* insertion mutant library (Tadege *et al*.., 2008; Noble Research Institute, LLC, http://medicago-mutant.noble.org/mutant/) were characterized for their flavonoid profiles. Plants homozygous for the insertions were identified by genomic PCR using *MtFBH-4* gene-specific primers and *tnt1* border primers (data not shown). Disruption of *MtFBH-4* in the homozygous mutant lines (R1 generation) was verified by RT-PCR (Fig. **4b**). Under the growth condition tested in this study, the homozygous *mtfbh-4* mutants did not show any noticeable phenotypic changes compared to the wild-type control plants.

Total metabolites were extracted from the aerial parts of 2-month-old plants and then subjected to HPLC-MS/MS analysis after removing *O*-conjugates by acid hydrolysis. As shown in Fig. **4c-d**, no significant difference in the amount of apigenin aglycone was observed between the homozygous mutants and the wild-type control plants, indicating that the first committed step in

flavone biosynthesis, the conversion of naringenin to apigenin, was not affected in both *mtfbh-4* mutants. However, complete depletion of tricin was observed in leaves and stems harvested from both homozygous mutant lines, in addition to a significant reduction of chrysoeriol when compared to metabolites extracted from wild-type plants (Fig. **4e-h**). Other potential flavone intermediates along the tricin pathway, including luteolin, selgin and tricetin, were not detected in both wild-type and mutant plants (data not shown). The substantial reduction in chrysoeriol content and the lack of tricin metabolites in *mtfbh-4* mutants suggest that MtFBH-4 catalyzes both 3′ and 5′-hydroxylations critical in completing the tricin pathway in the vegetative tissues of *M. truncatula*. Although MtFBH-5 (a canonical F3′H enzyme) is also expressed, its contribution to chrysoeriol accumulation is apparently limited. In fact, our kinetics analysis revealed its much lower substrate affinity towards apigenin compared to that of MtFBH-4; K_m of MtFBH-5 towards apigenin is about 10 times to that of MtFBH-4 (Supporting Information Fig. **S7a-b**).

To further investigate the types of tricin and chrysoeriol *O*-conjugates affected in the mutant lines, non-acid-hydrolyzed metabolites of both mutants were analyzed by HPLC-QTOF-MS and these flavone *O*-conjugates were identified by comparing their accurate mass and fragmentation pattern to those of flavone *O*-conjugates previously reported in *M. truncatula* (Fu & Wang, 2015; Jasinski *et al*., 2009; Kowalska *et al*., 2007; Staszków *et al*., 2011). Five tricin *O*-conjugates, including tricin-7-*O*-glucoside, tricin 7-*O*-glucuronide, tricin 7-*O*-glucuronyl-glucoside, tricin-7-*O*-di-glucuronoside and tricin 7-*O*-[2′-*O*-feruloyl-glucuronyl-(1-2)-glucuronide] could be readily detected in wild-type plants but were completely depleted in both mutant lines (Table **2**). Furthermore, the levels of chrysoeriol 7-*O*-glucuronide, chrysoeriol 7-*O*-glucuronyl-glucoside, chrysoeriol 7-*O*-diglucuronside and chrysoeriol 7-*O*-[2′-*O*-feruloyl-glucuronyl-(1-2)-glucuronide] were significantly reduced in the mutant lines compared with wild-type control (Table **2**).

Alfalfa orthologs of MtFBH-4 also demonstrate 5′-hydrxoylation activities

So far, our findings firmly establish that MtFBH-4, instead of canonical CYP75A F3′5′Hs, is a *bona fide* flavonoid hydroxylase that catalyzes 5′-hydroxylation leading to the production of extractable tricin metabolites in *M. truncatula*. As several alfalfa orthologs of MtFBH-4, including MsFBH-1, MsFBH-3, MsFBH-4 and MsFBH-10, harbor the same T492G amino acid polymorphism

at their SRS6 motifs (Table **1**) and are also phylogenetically distant from other *Medicago* CYP75B FBHs (Fig. **1**), they may also be involved in tricin biosynthesis.

Among this subgroup of alfalfa FBHs, only *MsFBH-4* and *MsFBH-10* were expressed in leaves and stems of wild-type plants whereas the expression of *MsFBH-1* and *MsFBH-3* could not be detected in both tissue types (Fig. **5a**). We then characterized the enzymatic activities of MsFBH-4 and MsFBH-10 under *in vitro* and *in planta* conditions by heterologous expression in yeast cells and *N. benthamiana* leaves, respectively, followed by incubation with various flavonoid substrates. HPLC-QTOF-MS analysis of the products from the *in vitro* enzyme assays revealed that both MsFBH-4 and MsFBH-10 convert naringenin, apigenin and kaempferol but not dihydrokaempferol into their corresponding 3′-hydroxylated flavonoids (Supporting Information Fig. **S3a-c** & **S8a)**. Furthermore, both MsFBH-4 and MsFBH-10 recognized 3′-methoxylated flavonoid substrates, including chrysoeriol, homoeriodictyol and isorhamnetin, and convert them into their corresponding 5′-hydroxylated flavonoids (Supporting Information Fig. **S8b-d**), but were incapable of modifying luteolin into tricetin or quercetin to myricetin (Supporting Information Fig. **S3d-e**). In addition, *N. benthamiana* leaves transiently expressing either *MsFBH-4* or *MsFBH-10* accumulated luteolin, chrysoeriol, selgin and tricin upon incubation with apigenin (Fig. **5b-f**), and selgin and tricin upon incubation with chrysoeriol (Supporting Information Fig. **S9a-c**). Overall, the two alfalfa orthologs of MtFBH-4, MsFBH-4 and MsFBH-10, sharing the same T492G amino acid polymorphism at the SRS6 motif, display enzyme activities highly resemble to those of MtFBH-4 in their abilities to introduce 3′-hydroxylations to a wide range of flavonoid substrates and more importantly, recognize several 3′-methoxylated flavonoid substrates and convert them into their corresponding 5′-hydroxylated forms.

Discussion

The distribution of tricin-type metabolites in dicots is relatively scattered when compared to in major monocot families including grasses (Poaceae) in which tricin exists as predominant soluble *O*-conjugates or abundantly incorporates into cell wall lignin (Lan *et al*., 2016; Li *et al*., 2016). Through studying *M. truncatula* and alfalfa CYP75B members, we identified a genus-specific subgroup of FBHs that catalyze 3′- and 5′-hydroxylations of the flavone B-ring. These reactions are essential for producing tricin derivatives, a unique metabolic trait that distinguishes these two *Medicago* species from other legumes and dicots. In fact, alfalfa remains the only dicot known to accumulate tricin-lignin in cell wall (Lan *et al*., 2015).

FBHs, consisting of CYP75A and CYP75B members, have been extensively studied for their involvement in anthocyanin production that determines flower coloration, a trait that has prominent impact on plant-pollinator interactions, interspecific competition and evolution (Holton *et al*., 1993; Tanaka & Brugliera, 2013). The divergence of CYP75A and CYP75B subfamilies is believed to occur before the emergence of seed plants (Seitz *et al*., 2006). The presence of one or more functional CYP75A F3′5′H in some angiosperm lineages allows the accumulation of 3′,5′-substituded (i.e., delphinidin-derived) anthocyanins, resulting in blue to violet flowers which may favor insect-mediated pollination (Harborne, 2014). Yet, 3′,5′-substituted flavonoids have patchy distributions in flowering plants due to repeated losses and/or inactivation of CYP75A genes during angiosperm evolution (Seitz *et al*., 2006; Tanaka & Brugliera, 2013). In this study, the absence of *CYP75A* expression in vegetative tissues of *M. truncatula* is apparently compensated by the independent recruitment of flavonoid 5′-hydroxylation activities in CYP75B MtFBH-4 for the production of tricin-derived metabolites. However, it is worth-noting that unlike CYP75A F3′5′H, MtFBH-4 failed to produce 3′,5′-dihydroxylated flavonoids from 3′-hydroxylated flavonoids such as luteolin and quercetin (Supporting Information Fig. **S3d-e**). Instead, MtFBH-4 accepts 3′-methoxylated flavonoids and produces 5′-hydroxyl-3′-methoxylated flavonoids (Fig. **2c-e** & Supporting Information Fig. **S5a-b**). Considering the substrate specificity of MtFBH-4 and the fact that chrysoeriol instead of tricetin was the differential flavonoid metabolite detected in the *mtfbh-4* mutants (Fig. **4e-f**), tricin biosynthesis in *M. truncatula* proceeds through chrysoeriol, i.e., via sequential hydroxylations and *O*-

methylations at the B-ring (Supporting Information Fig. **S1**), essentially recruiting the same biosynthetic route established previously in Poaceae (Lam *et al*., 2015; Lam *et al*., 2019a). In addition, MsFBH-4 and MsFBH-10 showed similar enzymatic activities to MtFBH-4 (Fig. **5b-e** & Supporting Information Fig. **S8a-d**), suggesting that the same enzymology is involved for tricin production in alfalfa, a closely related species of *M. truncatula*. In addition, our work further excluded the involvement of sequential 3′,5′-hydroxylations to produce tricetin from naringenin during tricin biosynthesis in *Medicago* spp. (Supporting Information Fig. **S1**).

In Poaceae (grasses), tricin metabolites are ubiquitously distributed and the CYP75B A3′H/C5′Hs required for producing tricin are also highly conserved amongst various family members, e.g., CYP75B4, CYP75B11 and CYP75B97 in rice, switchgrass and sorghum, respectively (Lam *et al*., 2015; Lam *et al*., 2019a). In comparison, the *Medicago* CYP75B FBHs with 5′-hydroxylation activities form a distinct clade in Fabaceae without sequences from other legume genera (Fig. **1**). In fact, the two identified *Trifolium pratense* CYP75B sequences do not cluster with these *Medicago* FBHs, although both genera belong to the same tribe Trifolieae. In addition, MtFBH-1 and MtFBH-2, which are immediately sister to the clade containing *Medicago* FBHs with 5′-hydroxylation activities (Fig. **1**), lack 5′-hydroxylation activities towards flavone substrates (Supporting Information Fig. **S4b-c** & **S6a-e**). Hence, functional divergence of CYP75B proteins for tricin production likely occurred during speciation of the *Medicago* genus. This distinguishing difference may account for the unique occurrence of tricin derivatives in *Medicago* spp., including tricin-lignins in alfalfa, but not in other members of Fabaceae such as soybean and lotus (Lan *et al*., 2016).

Consistently, multiple evolutions of F3′5′H from CYP75B FBHs had occurred in Asteraceae, giving rise to delphinidin-pigmented flowers in plants of the subfamilies Asteroideae and Cichoriodeae, and the genus *Echinops* of the subfamily Carduoideae (Seitz *et al*., 2015; Seitz *et al*., 2006). CYP75B enzymes isolated from Asteraceae species (e.g. *Callistephus chinensis*, *Echinops bannaticus, Cichorium intybus*) can convert flavanone (naringenin) and dihydroflavonol (dihydrokaempferol) into 3′,5′-dihydroxylated flavonoids, dihydrotricetin and dihydromyricetin, respectively, both serving as precursors for biosynthesis of delphinidin-derived pigments in petals and fruits (Seitz *et al*., 2006). However, each of these Asteraceae plants recruited F3′5′H activities

503 independently after the divergence from their common Asteraceae ancestor and consequently, some
504 Asteraceae plants, including those from the Mutisioideae subfamily and the Heliantheae tribe of
505 Asteroideae subfamily, do not accumulate delphinidin pigments (Seitz *et al*., 2015). Interestingly,
506 such gain-of-function in CYP75B members also occurred independently in *Billardiera heterophylla*,
507 which produces delphinidin-pigmented flowers but belongs to a different dicot family Pittosporaceae
508 (Seitz *et al*., 2015).

509 Since the three *Medicago* CYP75B FBHs (MtFBH-4, MsFBH-4 and MsFBH-10) recruited
510 their 5′-hydroxylation activities independently from Poaceae (grass) CYP75B A3′H/C5′Hs as well as
511 Pittosporaceae and Asteraceae CYP75B F3′5′Hs, we further examined their SRSs, key domains
512 determining activities of P450 enzymes, for amino acid polymorphisms which are commonly
513 associated with evolution and diversification of metabolic pathways (De Montellano, 2005; Dueholm
514 *et al*., 2015; Gotoh, 1992). According to protein sequence alignment, completely different amino acid
515 residues are incorporated into the 8th position of SRS6 (SRS6-8) of all three subgroups of CYP75B
516 proteins possessing 5′-hydroxylation activities (i.e., Asteraceae F3′5′Hs, Seitz *et al*., 2015; *Medicago*
517 FBHs, this work; Poaceae A3′H/C5′Hs, Lam *et al*., 2019), whereas a SRS6-8 Thr residue is strictly
518 conserved in all F3′H sequences examined (Table **1**). Also, these three unique subgroups of CYP75Bs
519 showed different substrate ranges for their 5′-hydroxylation activities. For example, Pittosporaceae
520 and Asteraceae F3′5′Hs, which harbor a Ser or Ala residue in SRS6-8, are able to catalyze sequential
521 3′,5′-hydroxylations of naringenin and dihydrokaempferol (Seitz *et al*., 2015). On the other hand,
522 *Medicago* FBHs with 5′-hydroxylation activities reported in this study harbor the smallest amino acid
523 residue, Gly, in SRS6-8 catalyze 3′-hydroxylation of a wide range of flavonoids but their 5′-
524 hydroxylation activities are restricted to 3′-methoxylated flavonoids (chrysoeriol, homoeriodictyol,
525 and isorhamnetin) which have a comparatively larger flavonoid B-ring than those of 3′-hydroxylated
526 flavonoids. However, when the Gly residue was artificially replaced by Thr in MtFBH-4, most of the
527 5′-hydroxylation activities were lost whilst the 3′-hydroxylation activity (towards apigenin) remained
528 intact (Fig. **2a-f**). Hence, the replacement of Thr residue by Gly residue in SRS6-8 may represent one
529 of key molecular events leading to the successful acquisition of 5′-hydroxylation activities in this
530 subgroup of *Medicago* FBHs. Interestingly, the Poaceae (grass) A3′H/C5′Hs, which harbor a Leu
531 residue in SRS6-8, only accept chrysoeriol for 5′-hydroxylation. Nonetheless, convergent evolution of

chrysoeriol 5′-hydroxylation had occurred at different phylogenetic levels for biosynthesis of tricin-derived metabolites, resulting in their restricted occurrences in *Medicago* spp. among Fabaceae plants but widespread distribution in Poaceae (grass).

It is also worth noting that the convergent evolution of hydroxylases and *O*-methyltransferases required for syringyl (S) lignins in angiosperm and *Selaginella* had also occurred at different phylogenetic levels (Weng *et al*., 2010). S lignins are ubiquitously found in angiosperms but in lycophytes, only some members in the order Selaginellales can synthesize S lignins (Weng *et al*., 2010). In particular, whereas 3- and 5-hydroxylations in the monolignol biosynthetic pathway in angiosperm are catalyzed separately by two different CYP enzymes, i.e., *p*-Coumaroyl ester 3-hydroxylase (C3′H, CYP98A) and ferulate 5-hydroxylase/coniferaldehyde 5-hydroxylase (F5H/CAld5H, CYP84), *Selaginella* lycophytes independently recruited a bifunctional phenylpropanoid *meta*-hydroxylase (CYP788A1) capable of catalyzing both 3- and 5-hydroxylation reactions required for S lignin biosynthesis, after separating from other lycophytes (Weng *et al*., 2008). Interestingly, this *Selaginella* hydroxylase could only catalyze 5-hydroxylation after 3-hydroxyl *O*-methylation (Weng *et al*., 2010), which is reminiscent of the catalytic abilities of grass A3′H/C5′Hs (Lam *et al*., 2019a) and the *Medicago* FBHs identified in this study.

Our metabolite analysis of *M. truncatula mtfbh-4* mutants indicates the direct involvement of MtFBH-4, and very likely MsFBH-4 and MsFBH-10, in modifying the flavone B-ring to produce tricin from apigenin in *Medicago* plants as no extractable tricin metabolites could be detected in the vegetative tissues of *mtfbh-4* mutant lines (Fig. **4g-h**). This notion is further consolidated by the successful reconstitution of tricin pathway in *N. benthamiana* leaves which transiently expressed the *Medicago FBH*s and fed with apigenin (Fig. **3a-e, 5b-f**) or chrysoeriol (Supporting Information Fig. **S5** & **S9**). Additionally, small amounts of extractable chrysoeriol metabolites were detected in the *mtfbh-4* mutants (Fig. **4e-f**), suggesting that at least one functional F3′H with relatively lower catalytic efficiency is present in *M. truncatula*. Considering the *in planta* expression pattern, enzyme activities and relatively lower binding affinity of the enzyme towards apigenin (Fig. **S4c-d**), this enzyme is likely to be MtFBH-5. However, this FBH obviously does not catalyze the 5′-hydroxylation of chrysoeriol to selgin as neither selgin nor tricin could be detected in *N. benthemiana*

leaves transiently expressing *MtFBH-5* incubated with apigenin (Fig. **5**) or in the *mtfbh-4* mutants.

Besides extractable metabolites, tricin may exist as an integrated component of cell wall lignin. As a natural lignin monomer, tricin undergoes oxidative co-polymerization with the conventional lignin monomers, i.e., monolignols, upon cell wall lignification in specific plant lineages (Lan *et al*., 2015; Lan *et al*., 2016; Li *et al*., 2016). Such tricin-incorporated lignin or tricin-lignin is abundantly and ubiquitously found in grasses including economically important food and biomass crops (Lan *et al*., 2016; Li *et al*., 2016). Recently, we demonstrated that manipulation of tricin biosynthesis in rice through *FNSII* or *A3′H/C5′H* disruption reduced lignin content in vegetative tissues and conferred them improved enzymatic saccharification efficiency, presenting a new strategy to improve grass biomass digestibility for biofuel and biomaterial production (Lam *et al*., 2019a; Lam *et al*., 2017). Interestingly, tricin-lignin has also been detected in alfalfa, a closely related species of *M. truncatula* (Lan *et al*., 2016). As the most cultivated forage legume worldwide, alfalfa is a major fodder crop for livestock. Genetic modifications of monolignol pathway genes in alfalfa had resulted in lignin reduction and improvement of cell wall digestibility, an important trait that affects the forage quality (Barros *et al*., 2019). In fact, HarvXtra is a commercial alfalfa product developed by down-regulation of a caffeoyl CoA 3-*O*-methyltransferase gene for lignin reduction (Barros *et al*., 2019). It is highly plausible that the tricin monomer comprising tricin-lignin in alfalfa is supplied through the functions of CYP75B FBHs, i.e., MsFBH-4 and MsFBH-10, identified in this study. It remains to be investigated whether targeting MsFBH-4, MsFBH-10 or other enzymes along tricin biosynthetic pathway can be strategic for reducing lignin content and enhancing cell wall digestibility and other forage qualities in alfalfa.

Altogether our results further strengthen the notion that there are multiple phylogenetic origins of flavonoid 5′-hydroxylation activities in the plant kingdom. In their course of evolution, *Medicago* plants possibly recruited flavonoid 5′-hydroxylation activities by incorporating amino acid polymorphisms (T492G) into redundant CYP75B FBHs along with other molecular events, allowing recognition and modifications of various substrates including chrysoeriol and the eventual completion of tricin biosynthesis pathway. Our findings further suggest that two unique alfalfa CYP75B FBHs with 5′-hydroxylation activities (MsFBH-4 and MsFBH-10) are potential candidates for improving

forage quality of alfalfa biomass via genetic manipulation of tricin-lignin. Moreover, our analysis may assist the prediction of 3′,5′-modified flavonoid accumulation among different plant species by examining the SRS6 of FBHs for unusual amino acid polymorphisms as observed in Poaceae (grass) CYP75B A3′H/C5′Hs (Lam *et al*., 2015; Lam *et al*. 2019a), Asteraceae CYP75B F3′5′Hs (Seitz *et al*., 2007; Seitz *et al*., 2015) and *Medicago* CYP75B FBHs (this work).

Acknowledgements

We thank the Noble Research Institute for providing the *M. truncatula* mutants which are jointly owned by the Centre National De La Recherche Scientifique and were created through research funded, in part, by a grant from the National Science Foundation, NSF-0703285. This work was supported by the Research Grants Council of Hong Kong (grant nos. GRF17113217 and GRF17126918) and the Japan Society for the Promotion of Science (JSPS; KAKENHI #16K14958 and #16H06198). ACWL acknowledges the Hong Kong PhD Fellowship Scheme. PYL acknowledges the JSPS fellowship programs (#17F17103).

Author contributions

ACWL, PYL and CL designed the experiments and analyzed data. ACWL, KHC and LW performed the experiments. ACWL, PYL, YT and CL wrote the manuscript with contributions of all other authors.

References

Akashi T, Fukuchi-Mizutani M, Aoki T, Ueyama Y, Yonekura-Sakakibara K, Tanaka Y, Kusumi T, Ayabe S. 1999. Molecular cloning and biochemical characterization of a novel cytochrome P450, flavone synthase II, that catalyzes direct conversion of flavanones to flavones. *Plant and Cell Physiology* **40**: 1182-1186.

Al-Fayez M, Cai H, Tunstall R, Steward WP, Gescher AJ. 2006. Differential modulation of cyclooxygenase-mediated prostaglandin production by the putative cancer chemopreventive flavonoids tricin, apigenin and quercetin. *Cancer Chemotherapy and Pharmacology* **58**: 816-

825.

Barros J, Temple S, Dixon RA. 2019. Development and commercialization of reduced lignin alfalfa. *Current Opinion in Biotechnology* **56**: 48-54.

Cai H, Hudson EA, Mann P, Verschoyle RD, Greaves P, Manson MM, Steward WP, Gescher AJ. 2004. Growth-inhibitory and cell cycle-arresting properties of the rice bran constituent tricin in human-derived breast cancer cells *in vitro* and in nude mice *in vivo*. *British Journal of Cancer* **91**: 1364-1371.

Cheng X, Wen J, Tadege M, Ratet P, Mysore KS. 2011. Reverse genetics in *Medicago truncatula* using *Tnt1* insertion mutants. In: Pereira A, ed. *Methods in Molecular Biology*. Totowa, NJ: Humana Press, **678**: 179-190.

Cummins I, Brazier-Hicks M, Stobiecki M, Franski R, Edwards R. 2006. Selective disruption of wheat secondary metabolism by herbicide safeners. *Phytochemistry* **67**: 1722-1730.

De Montellano PRO (ed). 2005. *Cytochrome P450: structure, mechanism, and biochemistry*. New York, US: Kluwer Academic/Plenum Publishers.

Del Río JC, Rencoret J, Prinsen P, Martínez AT, Ralph J, Gutiérrez A. 2012. Structural characterization of wheat straw lignin as revealed by analytical pyrolysis, 2D-NMR, and reductive cleavage methods. *Journal of Agricultural and Food Chemistry* **60**: 5922-5935.

Dixon RA, Pasinetti GM. 2010. Flavonoids and isoflavonoids: from plant biology to agriculture and neuroscience. *Plant Physiology* **154**: 453-457.

Dixon RA, Steele CL. 1999. Flavonoids and isoflavonoids - a gold mine for metabolic engineering. *Trends in Plant Science* **4**: 394-400.

Druka A, Kudrna D, Rostoks N, Brueggeman R, von Wettstein D, Kleinhofs A. 2003. *Chalcone isomerase* gene from rice (*Oryza sativa*) and barley (*Hordeum vulgare*): physical, genetic and mutation mapping. *Gene* **302**: 171-178.

Du Y, Chu H, Wang M, Chu IK, Lo C. 2010. Identification of flavone phytoalexins and a pathogen-inducible *flavone synthase II* gene (*SbFNSII*) in sorghum. *Journal of Experimental Botany* **61**: 983-994.

Duarte-Almeida JM, Negri G, Salatino A, de Carvalho JE, Lajolo FM. 2007. Antiproliferative and antioxidant activities of a tricin acylated glycoside from sugarcane (*Saccharum*

643 *officinarum*) juice. *Phytochemistry* **68**: 1165-1171.
644 **Dueholm B, Krieger C, Drew D, Olry A, Kamo T, Taboureau O, Weitzel C, Bourgaud F, Hehn**
645 **A, Simonsen HT. 2015.** Evolution of substrate recognition sites (SRSs) in cytochromes P450
646 from Apiaceae exemplified by the CYP71AJ subfamily. *BMC Evolutionary Biology* **15**: 122.
647 **Eloy NB, Voorend W, Lan W, Saleme ML, Cesarino I, Vanholme R, Smith RA, Goeminne G,**
648 **Pallidis A, Morreel K *et al*. 2017.** Silencing CHALCONE SYNTHASE in maize impedes the
649 incorporation of tricin into lignin and increases lignin content. *Plant Physiology* **173**: 998-
650 1016.
651 **Falcone Ferreyra ML, Emiliani J, Rodriguez EJ, Campos-Bermudez VA, Grotewold E, Casati**
652 **P. 2015.** The identification of maize and *Arabidopsis* Type I FLAVONE SYNTHASEs links
653 flavones with hormones and biotic interactions. *Plant Physiology* **169**: 1090-1107.
654 **Falcone Ferreyra ML, Rius SP, Casati P. 2012.** Flavonoids: biosynthesis, biological functions, and
655 biotechnological applications. *Frontiers in Plant Science* **3**: 222.
656 **Fu FY, Wang HL. 2015.** Metabolomics reveals consistency of the shoot system in *Medicago*
657 *truncatula* by HPLC-UV-ESI-MS/MS. *International Journal of Food Science and Technology*
658 **50**: 2183-2192.
659 **Gietz RD, Schiestl RH. 2007.** High-efficiency yeast transformation using the LiAc/SS carrier
660 DNA/PEG method. *Nature Protocols* **2**: 31-34.
661 **Gotoh O. 1992.** Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred
662 from comparative analyses of amino acid and coding nucleotide sequences. *Journal of*
663 *Biological Chemistry* **267**: 83-90.
664 **Han Y, Vimolmangkang S, Soria-Guerra RE, Rosales-Mendoza S, Zheng D, Lygin AV, Korban**
665 **SS. 2010.** Ectopic expression of apple F3´H genes contributes to anthocyanin accumulation in
666 the *Arabidopsis tt7* mutant grown under nitrogen stress. *Plant Physiology* **153**: 806-820.
667 **Harborne JB. 2014.** *Introduction to ecological biochemistry*. Cambridge, UK: Academic press.
668 **Harborne JB, Williams CA. 2000.** Advances in flavonoid research since 1992. *Phytochemistry* **55**:
669 481-504.
670 **Holsters M, de Waele D, Depicker A, Messens E, van Montagu M, Schell J. 1978.** Transfection
671 and transformation of *Agrobacterium tumefaciens*. *Molecular and General Genetics* **163**: 181-

187.

Holton TA, Brugliera F, Lester DR, Tanaka Y, Hyland CD, Menting JG, Lu CY, Farcy E, Stevenson TW, Cornish EC. 1993. Cloning and expression of cytochrome P450 genes controlling flower colour. *Nature* **366**: 276-279.

Jasinski M, Kachlicki P, Rodziewicz P, Figlerowicz M, Stobiecki M. 2009. Changes in the profile of flavonoid accumulation in *Medicago truncatula* leaves during infection with fungal pathogen *Phoma medicaginis*. *Plant Physiology and Biochemistry* **47**: 847-853.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**: 275-282.

Kim BG, Lee Y, Hur HG, Lim Y, Ahn JH. 2006. Flavonoid 3′-*O*-methyltransferase from rice: cDNA cloning, characterization and functional expression. *Phytochemistry* **67**: 387-394.

Kowalska I, Stochmal A, Kapusta I, Janda B, Pizza C, Piacente S, Oleszek W. 2007. Flavonoids from barrel medic (*Medicago truncatula*) aerial parts. *Journal of Agricultural and Food Chemistry* **55**: 2645-2652.

Krishnakumar V, Kim M, Rosen BD, Karamycheva S, Bidwell SL, Tang H, Town CD. 2015. MTGD: The *Medicago truncatula* genome database. *Plant and Cell Physiology* **56**: e1.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33**: 1870-1874.

Kwon YS, Kim EY, Kim WJ, Kim WK, Kim CM. 2002. Antioxidant constituents from *Setaria viridis*. *Archives of Pharmacal Research* **25**: 300-305.

Lam PY, Liu H, Lo C. 2015. Completion of tricin biosynthesis pathway in rice: cytochrome P450 75B4 is a unique chrysoeriol 5′-hydroxylase. *Plant Physiology* **168**: 1527-1536.

Lam PY, Lui ACW, Yamamura M, Wang L, Takeda Y, Suzuki S, Liu H, Zhu FY, Chen MX, Zhang J *et al.* 2019a. Recruitment of specific flavonoid B-ring hydroxylases for two independent biosynthesis pathways of flavone-derived metabolites in grasses. *New Phytologist* **223**: 204-219.

Lam PY, Tobimatsu Y, Matsumoto N, Suzuki S, Lan W, Takeda Y, Yamamura M, Sakamoto M, Ralph J, Lo C *et al.* 2019b. OsCAldOMT1 is a bifunctional *O*-methyltransferase involved in the biosynthesis of tricin-lignins in rice cell walls. *Scientific Reports* **9**: 11597.

Lam PY, Tobimatsu Y, Takeda Y, Suzuki S, Yamamura M, Umezawa T, Lo C. 2017. Disrupting flavone synthase II alters lignin and improves biomass digestibility. *Plant Physiology* **174**: 972-985.

Lam PY, Zhu FY, Chan WL, Liu H, Lo C. 2014. Cytochrome P450 93G1 is a flavone synthase II that channels flavanones to the biosynthesis of tricin *O*-linked conjugates in Rice. *Plant Physiology* **165**: 1315-1327.

Lan W, Lu F, Regner M, Zhu Y, Rencoret J, Ralph SA, Zakai UI, Morreel K, Boerjan W, Ralph J. 2015. Tricin, a flavonoid monomer in monocot lignification. *Plant Physiology* **167**: 1284-1295.

Lan W, Rencoret J, Lu F, Karlen SD, Smith BG, Harris PJ, Del Rio JC, Ralph J. 2016. Tricin-lignins: occurrence and quantitation of tricin in relation to phylogeny. *Plant Journal* **88**: 1046-1057.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al*. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Li M, Pu YQ, Yoo CG, Ragauskas AJ. 2016. The occurrence of tricin and its derivatives in plants. *Green Chemistry* **18**: 1439-1454.

Marles MA, Ray H, Gruber MY. 2003. New perspectives on proanthocyanidin biochemistry and molecular regulation. *Phytochemistry* **64**: 367-383.

Martens S, Mithofer A. 2005. Flavones and flavone synthases. *Phytochemistry* **66**: 2399-2407.

Mol JNM, Robbins MP, Dixon RA, Veltkamp E. 1985. Spontaneous and enzymic rearrangement of naringenin chalcone to flavanone. *Phytochemistry* **24**: 2267-2269.

Moreau C, Ambrose MJ, Turner L, Hill L, Ellis TH, Hofer JM. 2012. The *B* gene of pea encodes a defective flavonoid 3′,5′-hydroxylase, and confers pink flower color. *Plant Physiology* **159**: 759-768.

Peer WA, Murphy AS 2006. Flavonoids as signal molecules: targets of flavonoid action. In: Grotewold E, ed. *The science of flavonoids*. New York, US: Springer, 239-268.

Reddy AR, Scheffler B, Madhuri G, Srivastava MN, Kumar A, Sathyanarayanan PV, Nair S, Mohan M. 1996. Chalcone synthase in rice (*Oryza sativa* L.): detection of the CHS protein in

seedlings and molecular mapping of the *chs* locus. *Plant Molecular Biology* **32**: 735-743.

Sainsbury F, Thuenemann EC, Lomonossoff GP. 2009. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal* **7**: 682-693.

Schoenbohm C, Martens S, Eder C, Forkmann G, Weisshaar B. 2000. Identification of the *Arabidopsis thaliana* flavonoid 3′-hydroxylase gene and functional expression of the encoded P450 enzyme. *Biological Chemistry* **381**: 749-753.

Seitz C, Ameres S, Forkmann G. 2007. Identification of the molecular basis for the functional difference between flavonoid 3′-hydroxylase and flavonoid 3′,5′-hydroxylase. *FEBS Letter* **581**: 3429-3434.

Seitz C, Ameres S, Schlangen K, Forkmann G, Halbwirth H. 2015. Multiple evolution of flavonoid 3′,5′-hydroxylase. *Planta* **242**: 561-573.

Seitz C, Eder C, Deiml B, Kellner S, Martens S, Forkmann G. 2006. Cloning, functional identification and sequence analysis of *flavonoid 3′-hydroxylase* and *flavonoid 3′,5′-hydroxylase* cDNAs reveals independent evolution of flavonoid 3′,5′-hydroxylase in the Asteraceae family. *Plant Molecular Biology* **61**: 365-381.

Staszkow A, Swarcewicz B, Banasiak J, Muth D, Jasinski M, Stobiecki M. 2011. LC/MS profiling of flavonoid glycoconjugates isolated from hairy roots, suspension root cell cultures and seedling roots of *Medicago truncatula*. *Metabolomics* **7**: 604-613.

Tadege M, Wen J, He J, Tu H, Kwak Y, Eschstruth A, Cayrel A, Endre G, Zhao PX, Chabaud M *et al*. 2008. Large-scale insertional mutagenesis using the *Tnt1* retrotransposon in the model legume *Medicago truncatula*. *Plant Journal* **54**: 335-347.

Tanaka Y, Brugliera F. 2013. Flower colour and cytochromes P450. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**: 20120432.

Toda K, Yang DJ, Yamanaka N, Watanabe S, Harada K, Takahashi R. 2002. A single-base deletion in soybean *flavonoid 3′-hydroxylase* gene is associated with gray pubescence color. *Plant Molecular Biology* **50**: 187-196.

Weng JK, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C. 2010. Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte *Selaginella* and flowering

plants. *The Plant Cell* **22**: 1033-1045.

Weng JK, Li X, Stout J, Chapple C. 2008. Independent origins of syringyl lignin in vascular plants. *Proceedings of the National Academy of Sciences* **105:** 7887-7892.

Zhang J, Subramanian S, Stacey G, Yu O. 2009. Flavones and flavonols play distinct critical roles during nodulation of *Medicago truncatula* by *Sinorhizobium meliloti*. *Plant Journal* **57**: 171-183.

Zhang J, Subramanian S, Zhang Y, Yu O. 2007. Flavone synthases from *Medicago truncatula* are flavanone-2-hydroxylases and are important for nodulation. *Plant Physiology* **144**: 741-751.

Zhou JM, Gold ND, Martin VJ, Wollenweber E, Ibrahim RK. 2006. Sequential *O*-methylation of tricetin by a single gene product in wheat. *Biochimica et Biophysica Acta* **1760**: 1115-1124.

Zhou JM, Ibrahim RK. 2010. Tricin-a potential multifunctional nutraceutical. *Phytochemistry Reviews* **9**: 413-424.

Zhou JM, Seo YW, Ibrahim RK. 2009. Biochemical characterization of a putative wheat caffeic acid *O*-methyltransferase. *Plant Physiology and Biochemistry* **47**: 322-326.

Table 1. Multiple sequence alignment of the SRS6 domain of flavonoid B-ring hydroxylases from various plant species.

Protein	SRS6	CYP subfamily	Functional classification
Oyrza sativa CYP75B4	M D E A F T L **L** L	CYP75B	Poaceae A3´H/C5´Hs
Panicum virgatum CYP75B11	M E E A F T L **L** L		
Sorghum bicolor CYP75B97			
Zea mays ZmA3´H/C5´H			
Medicago sativa MsFBH-1/3/4/10	M D E A F G L **G** I	CYP75B	*Medicago* FBHs with 3´ & 5´-hydroxylation activities
Medicago truncatula MtFBH-4	M D E S F G L **G** I		
Medicago sativa MsFBH-5/8/9 & *Medicago truncatula* MtFBH-1/2	M D E G Y G L T L	CYP75B	Putative *Medicago* F3´Hs
Medicago sativa MsFBH-6/7	M D E G Y G F T L		
Medicago sativa MsFBH-2/11 & *Medicago truncatula* MtFBH-5	M D E A Y G L T L		
Arachis hypogaea AhF3´H-1/2	M D E A Y G L T L	CYP75B	Putative Fabaceae F3´Hs
Glycine max GmF3´H-3/4			
Lotus filicaulis LfF3´H			
Phaseolus vulgaris PvF3´H-1/2			
Trifolium pratense TpF3´H-1/2			
Glycine max GmF3´H-1/2	M D E T Y G I T L		
Callistephus chinensis CcF3´5´H	M D E A F G L **S** V	CYP75B	Asteraceae F3´5´Hs
Cichorium intybus CiF3´5´H	M E E A Y G L **A** L		
Echinops bannaticus EbF*3´*5´H	M E E A P G L **A** L		
Osteospermum hybrid OhF3´5´H	M E E E F G I **S** V		
Pericallis cruenta PcF3´5´H	M E E V F G I **S** L		
Arabidopsis thaliana AtF3´H	M E E S Y G L T L	CYP75B	Canonical F3´Hs

Oyrza sativa CYP75B3 *Sorghum bicolor* SbF3´H-1 *Zea mays* ZmF3´H-1 *Zea mays* ZmF3´H-2	M E E A Y G L T L		
Medicago sativa MsFBH-12/13 & Medicago truncatula MtFBH-6	M D E A F G L T L	CYP75A	Putative M*edicago* F3´5´Hs
Medicago sativa MsFBH-14 & Medicago truncatula MtFBH-7 &	M D E S F G L A L		
Glycine Max GmF3´5´H *Vitis vinifera* VvF3´5´H	M E E S F G L A L M D E A F G L A L	CYP75A	Canonical F3´5´Hs

The key amino acid residues determining the substrate recognition ability of flavonoid B-ring hydroxylases were underlined. At this amino acid position, a Thr or Ala residue can be found in canonical F3´Hs (CYP75B) and F3´5´Hs (CYP75A), respectively. In contrast, CYP75Bs with additional 5´-hydroxylation activities including those from Poaceae, Asteraceae and *Medicago* each harbors a different amino acid residue at the same position within SRS6 (bolded).

Table 2. HPLC-QTOF-MS analysis of chrysoeriol and tricin *O*-conjugates in *M. truncatula* wild-type and mutant leaves.

Compound annotated in *M. truncatula* leaves	*m/z*	MS/MS	Rt (min)	Peak area (x 10^4) per mg F.W.		
				WT	*mtf bh-4-a*	*mtfbh-4-b*
Chrysoeriol 7-*O*-glucuronide[4]	477.0990	301.0710	17.91	2.46 ± 0.36	0.17 ± 0.06*	0.13 ± 0.01*
Chrysoeriol 7-*O*-glucuronyl-glucoside[1]	639.1540	463.1250, 301.0710	16.54	8.74 ± 2.91	0.99 ± 0.52*	1.08 ± 0.38*
Chrysoeriol 7-*O*-diglucuronide[1,2,3]	653.1370	447.1050, 301.0710	16.52	16.33 ± 3.09	0.41 ± 0.19*	0.46 ± 0.12*
Chrysoeriol 7-*O*-[2'-*O*-feruloyl-glucuronyl-(1-2)-glucuronide][2]	829.1840	653.1700, 477.1040, 301.0710	17.56	2.62 ± 0.78	0.21 ± 0.10*	0.13 ± 0.01*
Tricin 7-*O*-glucoside[1]	493.1350	331.0820	17.96	2.54 ± 0.66	N.D.	N.D.
Tricin 7-*O*-glucuronide[3]	507.1140	331.0820	18.01	0.74 ± 0.27	N.D.	N.D.
Tricin 7-*O*-glucuronyl-glucoside[1]	669.1670	493.1370, 331.0820	16.76	1.67 ± 0.25	N.D.	N.D.
Tricin 7-*O*-diglucuronide[1,2,3]	683.1840	507.1140, 331.0820	16.67	3.50 ± 1.43	N.D.	N.D.
Tricin 7-*O*-[2'-*O*-feruloyl-glucuronyl-(1-2)-glucuronide][2,3]	859.1930	683.1570, 507.1150, 331.0820	17.70	0.84 ± 0.44	N.D.	N.D.

All chrysoeriol *O*-conjugates in *mtfbh-4-a* or *mtfbh-4-b* mutants were significantly reduced when compared to those of wild-type plants. Tricin *O*-conjugates could only be detected in *M. truncatula* wild-type plants but not in *mtfbh-4-a* and *mtfbh-4-b* mutants. Compound annotation was achieved via comparing the accurate mass and fragmentation pattern of individual peaks to those previously reported ([1]Fu & Wang, 2015; [2]Jasinski *et al*., 2009; [3]Kowalska *et al*., 2007; [4]Staszków *et al*., 2011). Data were expressed as mean ± 2SDs (n = 3;

Student's *t*-test, $*P < 0.05$). *m/z*, mass-to-charge ratio of the $[M + H]^+$ ions of flavone *O*-conjugates; MS/MS, major daughter ions generated from parental $[M + H]^+$ ions; Rt, retention time; F.W., fresh weight; WT, wild type; N.D., not detected.

774 **Figure and table legends**

775 **Figure 1.** Phylogenetic tree of CYP75B protein sequences.

776 The unrooted tree was constructed using the neighbor-joining method with 1000 bootstrapping
777 replications by MEGA7 (Kumar *et al*., 2016). Sequences from *M. sativa* were obtained by a BLASTP
778 search of the *M. sativa* protein sequence database (Noble Research Institute, LLC,
779 http://www.alfalfatoolbox.org/). Other sequences were retrieved via a BLASTP search of NCBI and
780 Phytozome (Phytozome V12.1, http://phytozome.jgi.doe.gov/) and their accession number were listed
781 in Supporting Information Table **S2**. All BLASTP searches were performed using GmF3´H as a query
782 (Toda *et al*., 2002). Cases of independent recruitment of flavonoid B-ring 5´-hydroxylase activities in
783 phylogenetically distant CYP75B proteins are highlighted in bold. Scale bar dentoes 0.05
784 substitutions per site.

785

786 **Figure 2.** Point mutation analysis of MtFBH-4.

787 (a) Nucleic acid (5´ to 3´) and amino acid (N terminal to C terminal) sequences at the SRS6 region of
788 wild-type and point mutant MtFBH-4 with differences bolded and underlined.

789 (b – e) HPLC-QTOF-MS detection of luteolin (b), selgin (c), 5´-hydroxyhomoeriodictyol (d) and
790 laricitrin (e) after incubation of recombinant MtFBH-4 WT or PM protein with apigenin, chrysoeriol,
791 homoeriodictyol and isorhamnetin, respectively.

792 (f) Substrate-to-production conversion rate of recombinant MtFBH-4 PM protein relative to WT
793 protein.

794 The identity of individual peaks from the spectra was confirmed by comparing the retention time and
795 fragmentation pattern with those of authentic standards, reaction products generated after incubation
796 of recombinant CYP75A1 with the same substrates or previous published spectrum (Lam *et al.,* 2015).
797 Quantification of flavonoids was performed by integrating the peak area of each flavonoid per 200 μg
798 microsomal protein used for enzyme assays. Data were expressed as mean ± 2SDs (n = 6, Student's t-
799 test, ***: $P < 0.001$). Mt, *M. truncatula*; WT, wild type; PM, point mutant. N.D., not detected.

Figure 3. *In planta* enzyme activities analyses of *M. truncatula* MtFBH-4 and MtFBH-5.

(a – e) HPLC-MS/MS detection of apigenin (a), luteolin (b), chrysoeriol (c), selgin (d) and tricin (e) after incubating *N. benthamiana* leaves transiently expressing *MtFBH-4* or *MtFBH-5* with apigenin.

(f) Biosynthetic route of tricin from apigenin in *N. benthamiana* leaves expressing *MtFBH-4*.

The identity of individual peaks from the spectra was confirmed by comparing the retention time and fragmentation pattern with those of authentic standards or previous published spectrum (Lam *et al.,* 2015). A3′H, apigenin 3′-hydroxylase; C5′H, chrysoeriol 5′-hydroxylase; FBH, flavonoid B-ring hydroxylase; FOMT, flavonoid *O*-methyltransferase; Mt, *M. truncatula*; Nb, *N. benthamiana*; XIC, extracted ion chromatography.

Figure 4. Analysis of *mtfbh-4* mutants.

(a) Gene structure of *MtFBH-4* and the *tnt1*-insertion site of its mutant lines, *mtfbh-4-a* and *mtfbh-4-b*.

(b) RT-PCR expression analysis of *MtFBH-4* in vegetative tissues of wild-type *Medicago truncatula* R108-1 and its corresponding *tnt1*-insertion mutants using primers CL4823+CL5054 for NF17189 (*mtfbh-4-a*) and primers CL5052+CL5053 for NF19367 (*mtfbh-4-b*) (Supporting Information Table **S1**). A housekeeping gene *Medicago truncatula polyubiquitin 3* (*MtUbi*; XM_003627103.1) was used as a positive control.

(c – h) HPLC-MS/MS detection of apigenin (c,d), chrysoeriol (e,f) and tricin (g,h) extracted from leaves and stems of *M. truncatula* R108-1 wild-type plants, *mtfbh-4-a* and *mtfbh-4-b*.

The identity of individual peaks from the spectra was confirmed by comparing the retention time and fragmentation pattern with those of authentic standards. WT, wild type; XIC, extracted ion chromatography.

Figure 5. *In planta* analyses of alfalfa MsFBH-4 and MsFBH-10.

(a) RT-PCR expression analysis of *MsFBH-1/3/4/10* in vegetative tissues of wild-type alfalfa plants.

825 A housekeeping gene *Medicago sativa retention in endoplasmic reticulum protein 1* (*MsRer1*;
826 Castonguay *et al*., 2015) was used as a positive control.

827 (b – f) HPLC-MS/MS detection of apigenin (b), luteolin (c), chrysoeriol (d), selgin (e; trace amount)
828 and tricin (f) after incubating *N. benthamiana* leaves transiently expressing *MsFBH-4* or *MsFBH-10*
829 with apigenin.

830 The identity of individual peaks from the spectra was confirmed by comparing the retention time and
831 fragmentation pattern with those of authentic standards or previous published spectrum (Lam *et al.,*
832 2015). FBH, flavonoid B-ring hydroxylase; Ms, *M. sativa*; Nb, *N. benthamiana*; XIC, extracted ion
833 chromatography.

834

835 **Table 1.** Multiple sequence alignment of the SRS6 domain of flavonoid B-ring hydroxylases from
836 various plant species.

837 The key amino acid residues determining the substrate recognition ability of flavonoid B-ring
838 hydroxylases were underlined. At this amino acid position, a Thr or Ala residue can be found in
839 canonical F3′Hs (CYP75B) and F3′5′Hs (CYP75A), respectively. In contrast, CYP75Bs with
840 additional 5′-hydroxylation activities including those from Poaceae, Asteraceae and *Medicago* each
841 harbors a different amino acid residue at the same position within SRS6 (bolded).

842

843

844 **Supporting information**

845 **Figure S1.** Proposed tricin biosynthetic pathway in the *Medicago* spp..
846 **Figure S2.** Expression analyses of all putative *Medicago truncatula* FBHs.
847 **Figure S3.** *In vitro* enzyme activities analyses of MtFBH-4, MsFBH-4 and MsFBH-10.
848 **Figure S4**. *In vitro* enzyme activities analyses of *M. truncatula* MtFBH-1, MtFBH-2 and MtFBH-5.
849 **Figure S5.** *In planta* enzyme activities analyses of *M. truncatula* MtFBH-4.
850 **Figure S6.** *In planta* enzyme activities analyses of *M. truncatula* MtFBH-1 and MtFBH-2.

Figure S7. Enzyme kinetics of *M*. *truncatula* MtFBH-4 and MtFBH-5 towards apigenin.

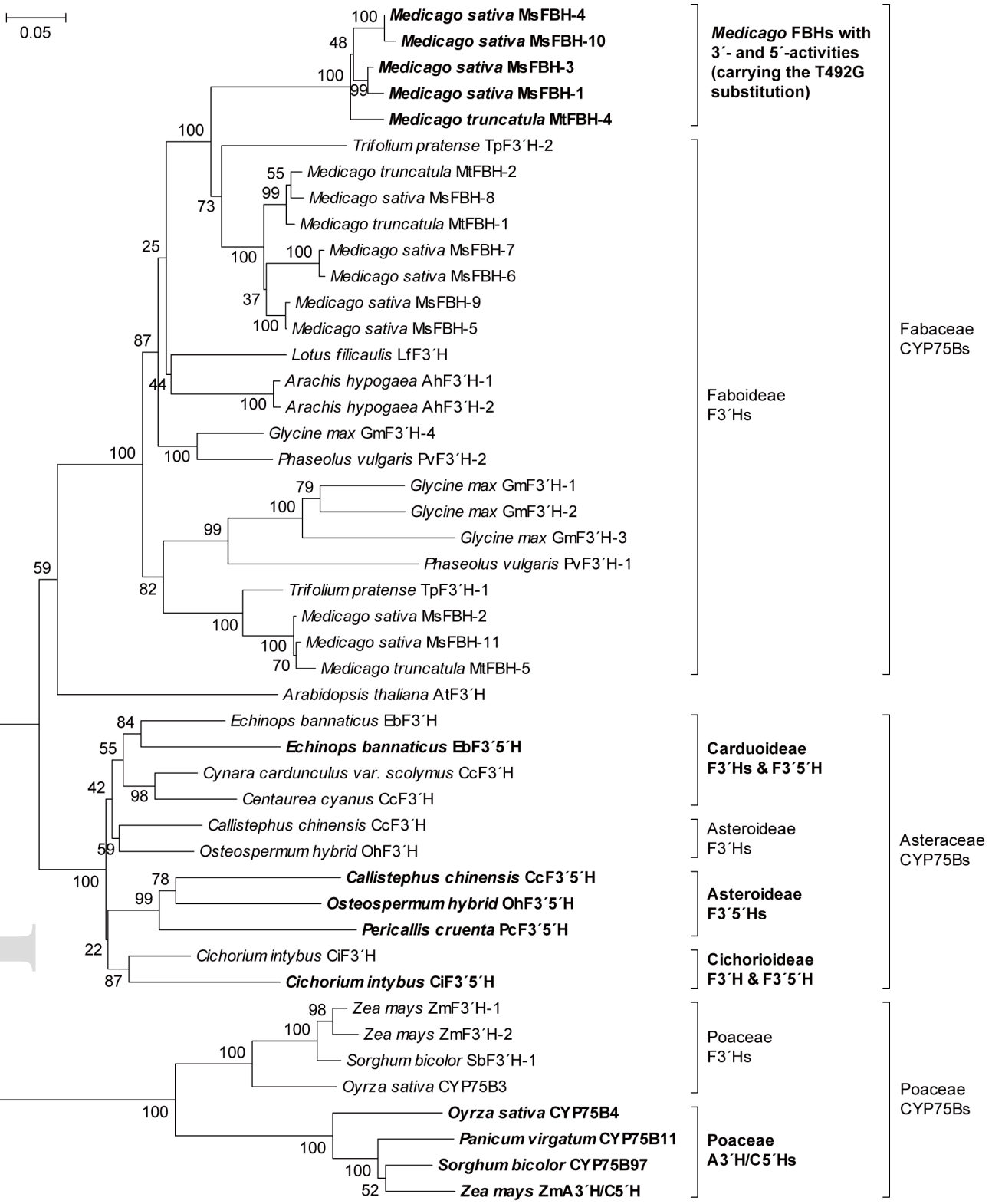
Figure S8. *In vitro* enzyme activities analyses of alfalfa MsFBH-4 and MsFBH-10.

Figure S9. *In planta* enzyme activities analyses of alfalfa MsFBH-4 and MsFBH-10.

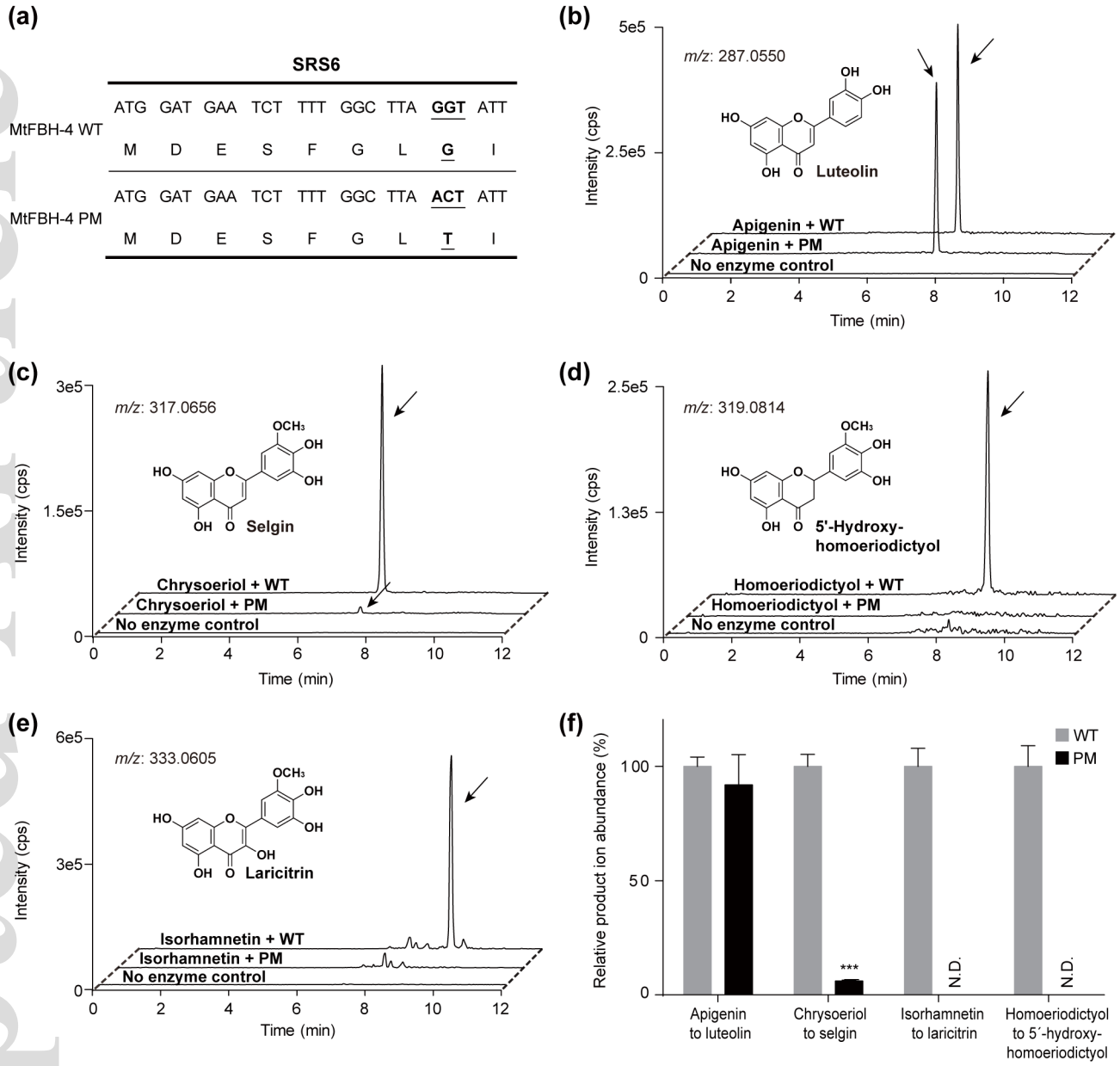
Table S1. Accession numbers or identifiers of putative *Medicago* flavonoid B-ring hydroxylases investigated in this study.

Table S2. Accession numbers or identifiers of flavonoid B-ring hydroxylases from CYP75B subfamily included in this study.

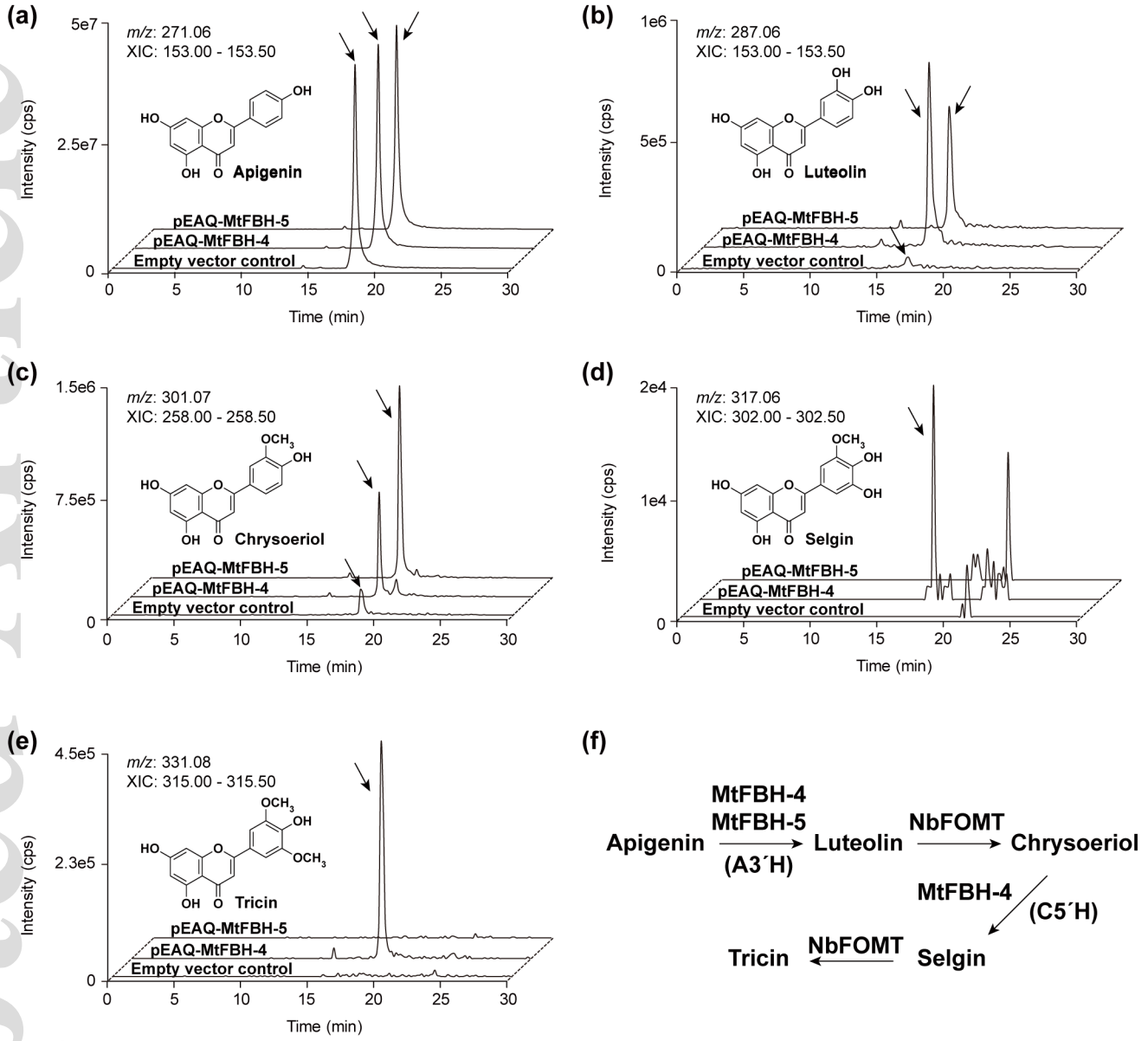
Table S3. Primers used for cloning, genotyping, RT-PCR and qRT-PCR experiments.



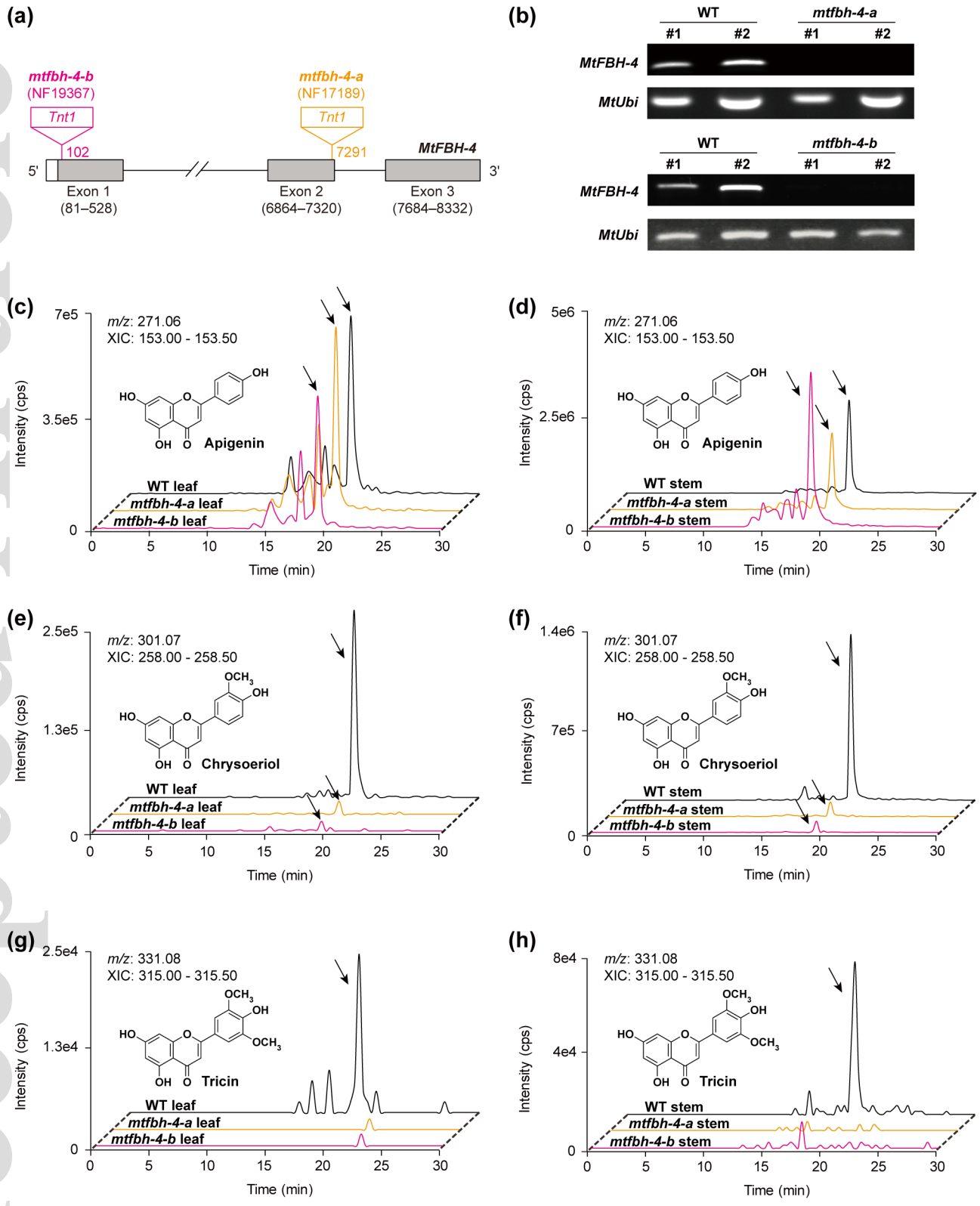
nph_16498_f1.tif



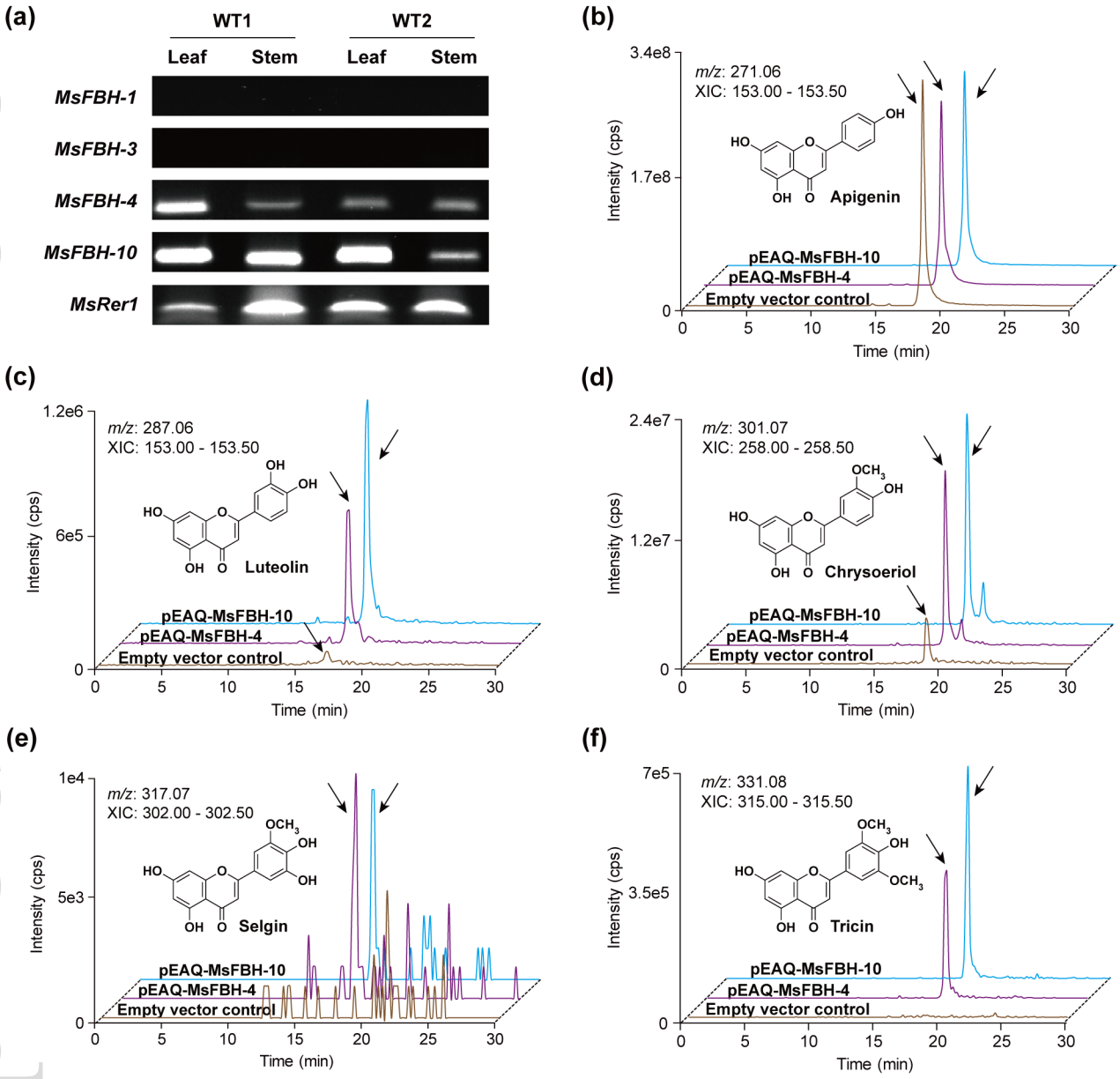
nph_16498_f2.tif



nph_16498_f3.tif



nph_16498_f4.tif



nph_16498_f5.tif