

Supplementary information

Identification and characterization of flavonoid biosynthetic enzyme genes in *Salvia miltiorrhiza*

Yuxing Deng ¹, Caili Li ¹, Heqing Li ^{1,2} and Shanfa Lu ^{1*}

¹ Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, No.151 Malianwa North Road, Haidian District, Beijing 100193, China; yuxingdeng2016@163.com (Y.D.); licaili390@163.com (C.L.); hqliaau@163.com (H.L.); sflu@implad.ac.cn (S.L.)

² College of Agronomy, Qingdao Agricultural University, No. 700 Changcheng Road, Chengyang District, Qingdao City, Shandong Province, 266109, China.

* Correspondence: sflu@implad.ac.cn (S.L.); Tel.: +86-10-57833366

Table S1. Sequence features of flavonoid metabolism pathway genes in *Arabidopsis thaliana* and other plants.

Gene name	Accession number	ORF (bp)	AA Len	MW (Da)	pI
<i>AtCHS</i>	AT5G13930	1188	395aa	43115.72	6.08
<i>AtCHI</i>	AT3G55120	741	246aa	26595.61	5.43
<i>AtF3H</i>	AT3G51240	1077	358aa	40275.72	5.29
<i>AtFLS1</i>	AT5G08640	1011	336aa	38281.77	5.63
<i>AtFLS2</i>	AT5G63580	753	250aa	28482.41	4.77
<i>AtFLS3</i>	AT5G63590	927	308aa	35451.53	5.13
<i>AtFLS4</i>	AT5G63595	840	279aa	31900.58	6.56
<i>AtFLS5</i>	AT5G63600	981	326aa	37092.85	4.66
<i>AtFLS6</i>	AT5G43935	882	293aa	33456.13	5.67
<i>AtF3'H</i>	AT5G07990	1542	513aa	56786.57	7.83
<i>AtDFR</i>	AT5G42800	1149	382aa	42774.96	5.43
<i>AtANS</i>	AT4G22880	1071	356aa	40396.42	5.23
<i>CsF3'5'H</i>	ABA40923.1	1533	510aa	57041.53	8.82
<i>GmF3'5'H</i>	AAM51564.1	1527	508aa	57185.75	9.11

Table S2. Primers used for full-length coding region cloning.

Gene name	Primer name	Sequence (5' to 3')
<i>SmCHS1</i>	CHS1-F	AACAAGAATGGTGACCGTGGA
	CHS1-R	TCAATTGATGTTGACGCTGTGCAG
<i>SmCHS2</i>	CHS2-F	CCAATGTGCAAGAATCCAATTTAGAG
	CHS2 -R	GTTGCACACCATGTTAAATATAGC
<i>SmCHS3</i>	CHS3 -F	ATGGCAAGCGTGGAGGAGATC
	CHS3 -R	TTATAAATCAATCGGCATGCTG
<i>SmCHS4</i>	CHS4 -F	ATGGCGACCGTGGAGGAGATC
	CHS4 -R	TTAATTAATCGGCATACTACGTAGC
<i>SmCHS5</i>	CHS5 -F	ATGGCAAGCGTGGAGGAGATC
	CHS5 -R	TTATAATTCAATCGGAATGCTGCGC
<i>SmCHS6</i>	CHS6 -F	TGTAATGTCGACAAAGCAGCAG
	CHS6 -R	GATAGTTAATGGAGGAGGCTTCGC
<i>SmCHS7</i>	CHS7-F	CAAACCCCGCCTAACAACAACG
	CHS7-R	CGTTATATTTAAATCGGCATGCTGTG
<i>SmCHS8</i>	CHS8-F	CAAACCCCGCCTAACAACAACG
	CHS8-R	GAACACAGACTAGAGGCCCGAGAC
<i>SmCHI1</i>	CHI1-F	CCACCTTCATCTTCATCTTCACC
	CHI1 -R	CATGTTTCTTGATTCAATTGTGCCTC

Table S2. *Cont.*

<i>SmCHI2</i>	CHI2-F CHI2-R	CTTCATCTTCGCCTTCACCTTC CATGTTTCTTGATTCAATTGTGCCTC
<i>SmCHI3</i>	CHI3-F CHI3-R	GAGGCTGTGATGGTGGATGAAGTTC TTATTTGGATAACTCGGCGGAGAGAG
<i>SmCHI4</i>	CHI4-F CHI4 -R	CCATTGTTAACCACCTTCATCTTCAC CATGTTTCTTGATTCAATTGTGCCTCA
<i>SmF3H1</i>	F3H1-F F3H1-R	CATGGCTGAAATGAAAATATCTCC TGGTGGAATCAACCTTAATTTAGAA
<i>SmF3H2</i>	F3H2-F F3H2-R	ATGGCTGCACAAATGAAAATCTG TAATAAAGGGTTGTTTGGCTGGC
<i>SmF3'5'H</i>	F3'5'H-F F3'5'H-R	CAACATATAGTAGGCATGCAAGGTG ACCTAGCTAGCTGCATAACAATGTG
<i>SmF3'H1</i>	F3'H1-F F3'H1-R	ATATATTCACGTTGGCATGCA GCAAACACTTAAGCATGGTAAAC
<i>SmF3'H2</i>	F3'H 2-F F3'H 2-R	ACACACACATCACATATATGCAATG TCAGATTTGGGCTTGATAAACATGG
<i>SmF3'H3</i>	F3'H 3-F F3'H 3-R	ATGCAGAGCTTCCACGAATTTCTAT GCAACGGTACATTCATTCATCAAC
<i>SmF3'H4</i>	F3'H 4-F F3'H 4-R	CCTTGTCAAATTAGTTTCGCTTCAG CCAGTCATCACTCATACAAACTAGG
<i>SmF3'H5</i>	F3'H 5-F F3'H 5-R	CCTCTCAAACACAACCTCTCATCG CAAACCGAGGCATAAAATTCACATAG
<i>SmF3'H6</i>	F3'H 6-F F3'H 6-R	CTCTCTCCAAATTTAATTAGCAATGG GCAGAGCCATAAATTCACATAG
<i>SmFLS1</i>	FLS1-F FLS1-R	CTATACCACCACCATGTAAGATCAC CTCAAATCAGGGACAGTGGTTAT
<i>SmFLS2</i>	FLS2-F FLS2-R	AAATGGAGGTTGAGAGAGTGC TATTTGTTTGGTGGGTAGAGTGA
<i>SmFNS II</i>	FLS II -F FLS II -R	GCGAGTTGTACTGATATGGAGCTAG ACAGTAGCTAAACCGAATCAGGC
<i>SmDFR</i>	DFR-F DFR-R	AATGCCGCTAGAAACCACCCCTC TCAATCACATTTGCCATTCTCGCCT
<i>SmANS</i>	ANS-F ANS-R	CCATGCATTTCACCCCTAACATAC TACAACTCACACAATCAACTCGA

Table S3. Primers used for qRT-PCR.

Gene name	Primer name	Sequence (5' to 3')
<i>SmUBQ10</i>	UBQ ₁₀ -RT-F	AGATGGGCGGACACTTGCTGATTA
	UBQ ₁₀ -RT-R	ACTCTCCACCTCCAAAGTGATGGT
<i>SmCHS1</i>	CHS1-RT-F	CAAGCGCATGTGTGAAAAGTCGA
	CHS1-RT-R	TGCAGAAGACGAGGTGGGTGA
<i>SmCHS2</i>	CHS2-RT-F	ATGGGAATCGGTGAATTGTGTAA
	CHS2-RT-R	GCAAAGTCGGACAAGCTTCTG
<i>SmCHS3</i>	CHS3 –RT-F	GGCTTCTTATCGATCTCGTCAGA
	CHS3 –RT-R	CCAACCTTCGCCTCCACCTGAT
<i>SmCHS4</i>	CHS4 –RT-F	CATTTAAGCGCATGTGTGAGAGC
	CHS4 –RT-R	CATGTGGATGCTGCCGTTGGT
<i>SmCHS5</i>	CHS5 –RT-F	GTCTGCTGCGAGTCCACC
	CHS5 –RT-R	GATACCTTTTGCGAGATCTATGAGA
<i>SmCHS6</i>	CHS6 –RT-F	ATGTCGACAAAGCAGCAGCGC
	CHS6 –RT-R	TCGTCGTTTTGCAGAGGCGTTG
<i>SmCHS7</i>	CHS7-RT-F	CAAGCATCGCCGACTACGA
	CHS7-RT-R	CAAGTCGGCGGAAGTGTT
<i>SmCHS8</i>	CHS8-RT-F	CGTTTCCTGATTACTACTTCCGTGT
	CHS8-RT-R	CGGGATGTTGGGTTTTCTTTGA
<i>SmCHI1</i>	CHI1-RT-F	TGCCGATGAACTGGCCGATTCCA
	CHI1-RT-R	CCGTTAGTGGCAGAATCATAGTC
<i>SmCHI2</i>	CHI2-RT-F	AGTCTTCAAGGATGAAAGCTTCTC
	CHI2-RT-R	CCAAGCTCTGCCTCGTGTA
<i>SmCHI3</i>	CHI3-RT-F	GCTCCTACCGTCCAGATTGTATT
	CHI3-RT-R	TTATTTGGATAACTCGGCCGA
<i>SmCHI4</i>	CHI4-RT-F	TTCCCACCCACAGCGAAGCCACG
	CHI4-RT-R	AATGAAGTCGGCGGAATCGG
<i>SmF3H1</i>	F3H1-RT-F	CAATGGGAGGTTCAAGAATGCAG
	F3H1-RT-R	GAAAAGAGGTTTGTGTTTGTTGGCG
<i>SmF3H2</i>	F3H2-RT-F	GACCCGGGATGACGGCAACACAT
	F3H2-RT-R	GAACCTCCCATTTGCTTACATAAT
<i>SmF3'5'H</i>	F3'5'H-RT-F	ACTGAACTTGTTTACTGCAGGGAC
	F3'5'H-RT-R	GGTATGCTAGTTTCGGAATGTCTG
<i>SmF3'H1</i>	F3'H1-RT-F	TTAGACAGGAGGAAGCGGGGATT
	F3'H1-RT-R	ACTCCTCCGCCTTCTCGTCTCCA

Table S3. *Cont.*

<i>SmF3'H2</i>	F3'H 2-RT-F	TCCGCCATGTTAGAGAGGAAGAAGT
	F3'H 2-RT-R	TCCACCATCGCCTTCAACTCTT
<i>SmF3'H3</i>	F3'H 3-RT-F	CCTTGGCTTAGTTTCCTCGATTTG
	F3'H 3-RT-R	TCAGTACCACCAGTTAGCAAGTCCT
<i>SmF3'H4</i>	F3'H 4-RT-F	CATGGATATGGTAGTAGGAGGAACG
	F3'H 4-RT-R	GCAGTTTGTGATGTGGGATTCTT
<i>SmF3'H5</i>	F3'H 5-RT-F	AGACTGTAAGGTGGATGGGTATGA
	F3'H 5-RT-R	TCGAAATCTTGTCCCTTCACAT
<i>SmF3'H6</i>	F3'H 6-RT-F	ACATGACTGTAAGGTGGATGGGT
	F3'H 6-RT-R	TCGAAATCCTGTCCCTTCACAT
<i>SmFLS1</i>	FLS1-RT-F	CCTGATTATAGGGAGGTTACTGAAG
	FLS1-RT-R	ACCCAAGAACCAAGTCCGG
<i>SmFLS2</i>	FLS2-RT-F	GGCGACCAAATTGAGATACTAAG
	FLS2-RT-R	TCACAAGCTTTGGAATTGGC
<i>SmFNS II</i>	FNS II -RT-F	ATTCTGGATTTTTTTACCGCCG
	FNS II -RT-R	CTCTCTGATCCCACCACTTTTGC
<i>SmDFR</i>	DFR-RT-F	GATGATCCCGAGAATGAAGTGATC
	DFR-RT-R	AGACACAAAGTACATCCATCCAGTC
<i>SmANS</i>	ANS-RT-F	CTGATTATATACCGGCTACATG
	ANS-RT-R	AATCCGCCGACTTCTTTCTC

Supplementary Table S4. The proteins used in the phylogenetic tree of CHS.

Protein name	Plant Species	Accession number
OsCHS1	<i>Oryza sativa</i>	A2ZEX7.1
OsCHS2	<i>Oryza sativa</i>	BAA19186.2
ZmCHS(C2)	<i>Zea mays</i>	CAA42764.1
ZmCHS(Whp)	<i>Zea mays</i>	CAA42763.1
ZmCHSL	<i>Zea mays</i>	NP_001149508.1
GhyCHS1	<i>Gerbera hybrida</i>	CAA86218.1
GhyCHS3	<i>Gerbera hybrida</i>	CAA86220.1
GhyCHS4	<i>Gerbera hybrida</i>	CAP20328.1
AtCHS	<i>Arabidopsis thaliana</i>	NP_196897.1
AtPKS-B	<i>Arabidopsis thaliana</i>	Q8LDM2.1
AtPKS-A	<i>Arabidopsis thaliana</i>	O23674.1
VvCHS1	<i>Vitis vinifera</i>	BAA31259.1
VvCHS2	<i>Vitis vinifera</i>	BAB84112.1
PcCHS	<i>Petroselinum crispum</i>	P16107.1
RgCHS1	<i>Ruta graveolens</i>	Q9FSB9.1
RgCHS2	<i>Ruta graveolens</i>	Q9FSB8.1
RgCHS3	<i>Ruta graveolens</i>	Q9FSB7.1
SvCHS	<i>Scutellaria viscidula</i>	ACC68839.1
SbaCHS1	<i>Scutellaria baicalensis</i>	AMW91735.1
SbaCHS2	<i>Scutellaria baicalensis</i>	AMW91736.1
AmCHS	<i>Antirrhinum majus</i>	CAA27338.1
ArCHS	<i>Agastache rugosa</i>	AFL72079.1
SsCHS	<i>Solenostemon scutellarioides</i>	ABP57071.1
PfCHS	<i>Perilla frutescens</i>	BAA19548.1
IpCHSA	<i>Ipomoea purpurea</i>	AAB02620.1
IpCHSB	<i>Ipomoea purpurea</i>	AAC49030.1
IpCHSC	<i>Ipomoea purpurea</i>	AAC49031.1
IpCHSD	<i>Ipomoea purpurea</i>	ABW69675.1
IpCHSE	<i>Ipomoea purpurea</i>	BAA87337.1
IpCHSFL1	<i>Ipomoea purpurea</i>	AAB41103.1
PhCHSA	<i>Petunia hybrida</i>	CAA32731.1
PhCHSB	<i>Petunia hybrida</i>	CAA32732.1
PhCHSD	<i>Petunia hybrida</i>	CAA32733.1
PhCHSF	<i>Petunia hybrida</i>	CAA32734.1
PhCHSG	<i>Petunia hybrida</i>	CAA32735.1
PhCHSH	<i>Petunia hybrida</i>	CAA32736.1
PhCHSJ	<i>Petunia hybrida</i>	CAA32737.1
SlCHS1	<i>Solanum lycopersicum</i>	CAA38980.1
SlCHS2	<i>Solanum lycopersicum</i>	CAA38981.1
MdCHS1	<i>Malus domestica</i>	AAY45748.1
MdCHS2	<i>Malus domestica</i>	AGE84303.1

Table S4. Cont.

MdCHS3	<i>Malus domestica</i>	ACJ54531
NnCHS	<i>Nelumbo nucifera</i>	ADD74168.1
GmCHS1	<i>Glycine max</i>	NP_001337038.1
GmCHS2	<i>Glycine max</i>	CAA46590.1
GmCHS3	<i>Glycine max</i>	NP_001276296.1
GmCHS4	<i>Glycine max</i>	CAA36317.1
GmCHS5	<i>Glycine max</i>	AAB01004.1
GmCHS6	<i>Glycine max</i>	AAA33951.1
GmCHS7	<i>Glycine max</i>	AAA33950.1
GmCHS8	<i>Glycine max</i>	NP_001304585.2
GmCHS9	<i>Glycine max</i>	ABQ63059.1
MsCHS1	<i>Medicago sativa</i>	AAA02823.1
MsCHS2	<i>Medicago sativa</i>	AAA02824.1
MsCHS3	<i>Medicago sativa</i>	CAA48226.1
MsCHS4	<i>Medicago sativa</i>	AAA02825.1
MsCHS8	<i>Medicago sativa</i>	AAA02826.1
MsCHS9	<i>Medicago sativa</i>	AAA02827.1
PnCHS	<i>Psilotum nudum</i>	BAA87922.1
EaCHS	<i>Equisetum arvense</i>	BAA89501.1
GbCHS	<i>Ginkgo biloba</i>	AAY52458.1
PstCHS	<i>Pinus strobus</i>	CAA06077.1
PsyCHS	<i>Pinus sylvestris</i>	CAA43166.1
PrCHS	<i>Pinus radiata</i>	AGY80771.1
PrCHSL	<i>Pinus radiata</i>	AAB80804.1
PhaCHS	<i>Phalaenopsis hybrida</i>	AAV70116.1
PpaCHS	<i>Physcomitrella patens</i>	ABB84527.1
NsCHSL	<i>Nicotiana sylvestris</i>	CAA74847.1

Table S5. The proteins used in the phylogenetic tree of CHI

Protein name	Plant Species	Accession number
AtCHI	<i>Arabidopsis thaliana</i>	At3g55120
AtCHIL	<i>Arabidopsis thaliana</i>	AT5g05270
AtFAP1	<i>Arabidopsis thaliana</i>	At3g63170
AtFAP2	<i>Arabidopsis thaliana</i>	At2g26310
AtFAP3	<i>Arabidopsis thaliana</i>	At1g53520
LjCHI1	<i>Lotus japonicus</i>	BAC53983.1
LjCHI2	<i>Lotus japonicus</i>	BAC53984.1
LjCHI3	<i>Lotus japonicus</i>	BAC54038.1
LjCHI4	<i>Lotus japonicus</i>	BAC53984.1
AhCHI I	<i>Arachis hypogaea</i>	AFP33452.1
AhCHI II	<i>Arachis hypogaea</i>	AJF45950.1
GmCHI1A	<i>Glycine max</i>	Glyma.20G241500.1
GmCHI1B1	<i>Glycine max</i>	Glyma.20G241600.1
GmCHI1B2	<i>Glycine max</i>	Glyma.10G292200.1
GmCHI2	<i>Glycine max</i>	Glyma.20G241700.1
GmCHI3A1	<i>Glycine max</i>	Glyma.13G262500.1
GmCHI3A2	<i>Glycine max</i>	Glyma.15G242900.1
GmCHI3B1	<i>Glycine max</i>	Glyma.03G154600.1
GmCHI3B2	<i>Glycine max</i>	Glyma.19G156900.1
GmCHI3C1	<i>Glycine max</i>	Glyma.14G098100.1
GmCHI3C2	<i>Glycine max</i>	Glyma.17G226600.1
GmCHI4A	<i>Glycine max</i>	Glyma.06G143000.1
GmCHI4B	<i>Glycine max</i>	Glyma.04G222400.1
VvCHI1	<i>Vitis vinifera</i>	P51117.1
VvCHI2	<i>Vitis vinifera</i>	A5ANT9.1
PfCHI	<i>Perilla frutescens</i>	BAG14301.1
SbaCHI	<i>Scutellaria baicalensis</i>	ADQ13184.1
ArCHI	<i>Agastache rugosa</i>	AFL72080.1
InCHIL	<i>Ipomoea nil</i>	BAO58578.1
GhCHI	<i>Gossypium hirsutum</i>	ABM64798.1
SmeCHI	<i>Saussurea medusa</i>	AAM48130.1
CsCHI	<i>Camellia sinensis</i>	AAZ17563.2
DcCHI	<i>Dianthus caryophyllus</i>	CAA91931.1
PcCHI	<i>Pyrus communis</i>	ABQ08639.1
IpCHI	<i>Ipomoea purpurea</i>	ABW69677.1
PhCHIA	<i>Petunia hybrida</i>	AAF60296.1
PhCHIB	<i>Petunia hybrida</i>	P11651.1
RsCHI	<i>Raphanus sativus</i>	O22651.1
NtCHI	<i>Nicotiana tabacum</i>	Q33DL3.1
ZmCHI	<i>Zea mays</i>	Q08704.1
HvCHI	<i>Hordeum vulgare</i>	AAM13449.1

Table S5. Cont.

OsCHI	<i>Oryza sativa</i>	AAM13448.1
SICHI1	<i>Solanum lycopersicum</i>	Solyc05g010320.2.1
SICHI2	<i>Solanum lycopersicum</i>	Solyc05g010310.2.1
SICHI3	<i>Solanum lycopersicum</i>	Solyc05g052240.2.1
SICHI4	<i>Solanum lycopersicum</i>	Solyc07g062030.2.1
SICHI5	<i>Solanum lycopersicum</i>	Solyc08g061480.2.1
SICHI6	<i>Solanum lycopersicum</i>	Solyc02g067870.2.1
SICHI7	<i>Solanum lycopersicum</i>	Solyc06g084260.2.1
MsCHI	<i>Medicago sativa</i>	P28012.1
PvCHI	<i>Phaseolus vulgaris</i>	P14298.2
PICHI	<i>Pueraria lobata</i>	Q43056.1
LaCHIL1	<i>Lupinus angustifolius</i>	CCM80407.1
LaCHIL2	<i>Lupinus angustifolius</i>	CCM80406.1

Table S6. The proteins used in the phylogenetic tree of FNSII, F3'5'H and F3'H

Protein name	Plant Species	Accession number
PfFNSII	<i>Perilla frutescens</i>	BAB59004.1
SbaFNSII-1	<i>Scutellaria baicalensis</i>	AMW91728.1
SbaFNSII-2	<i>Scutellaria baicalensis</i>	AMW91729.1
ThFNSII	<i>Torenia hybrida</i>	BAA84072
AmFNSII	<i>Antirrhinum majus</i>	BAA84071.1
GtFNSII	<i>Gentiana triflora</i>	BAD91809.1
GmFNSII	<i>Glycine max</i>	ACV65037.1
LeFNSII	<i>Lobelia erinus</i>	BAF49323.1
CcFNSII	<i>Callistephus chinensis</i>	AAF04115.1
PoFNSII	<i>Pilosella officinarum</i>	ACB56919.1
GhyFNSII	<i>Gerbera hybrida</i>	AAD39549.1
CsFNSII	<i>Camellia sinensis</i>	ACH99109.1
GeFNSII	<i>Glycyrrhiza echinata</i>	P93149.2
MtFNSII	<i>Medicago truncatula</i>	ABC86159.1
SbFNSII	<i>Sorghum bicolor</i>	XP_002461286.1
OsFNSII	<i>Oryza sativa</i>	BAG94859.1
ZmFNSII	<i>Zea mays</i>	XP_008663013.1
GmF3'H	<i>Glycine max</i>	BAD97828.1
MtF3'H	<i>Medicago truncatula</i>	XP_003598945.1
MdF3'H	<i>Malus domestica</i>	ACR14867.1
GhF3'H	<i>Gossypium hirsutum</i>	ADP24159.1
BnF3'H	<i>Brassica napus</i>	ABC58722.1
AtF3'H	<i>Arabidopsis thaliana</i>	Q9SD85.1
ThF3'H	<i>Torenia hybrida</i>	BAB87838.1
AmF3'H	<i>Antirrhinum majus</i>	ABB53383.1
PfF3'H	<i>Perilla frutescens</i>	BAB59005.1
GtF3'H	<i>Gentiana triflora</i>	BAD91808
IbF3'H	<i>Ipomoea batatas</i>	AEH42499.1
InF3'H	<i>Ipomoea nil</i>	BAD00190
IpF3'H	<i>Ipomoea purpurea</i>	AAR00229.1
SIF3'H	<i>Solanum lycopersicum</i>	NP_001289844.1
PhF3'H (Ht1)	<i>Petunia hybrida</i>	Q9SBQ9.1
LeF3'H	<i>Lobelia erinus</i>	BAF49324.1
VvF3'H1	<i>Vitis vinifera</i>	BAE47003.1
VvF3'H2	<i>Vitis vinifera</i>	BAE47003.2
VvF3'H3	<i>Vitis vinifera</i>	BAE47003.3
VvF3'H4	<i>Vitis vinifera</i>	BAE47003.4
CsF3'H1	<i>Camellia sinensis</i>	AKJ86992
CsF3'H2	<i>Camellia sinensis</i>	AKM12329.1
CsF3'H3	<i>Camellia sinensis</i>	AKM12330.1

Table S6. Cont.

CnF3'H	<i>Camellia nitidissima</i>	ADZ28515.1
OhF3'H	<i>Osteospermum hybrida</i>	ABB29899.1
GhyF3'H	<i>Gerbera hybrida</i>	ABA64468
EsF3'H	<i>Epimedium sagittatum</i>	ADE80941.1
SbF3'H1	<i>Sorghum bicolor</i>	ABG54319.1
SbF3'H2	<i>Sorghum bicolor</i>	ABG54320.1
SbF3'H3	<i>Sorghum bicolor</i>	ABG54321.1
ZmF3'H	<i>Zea mays</i>	AEF33624.1
GbF3'H-like	<i>Ginkgo biloba</i>	AJO67233.1
SiF3'H-like	<i>Sesamum indicum</i>	XP_011095827.1
PhyF3'5'H	<i>Phalaenopsis hybrida</i>	AAZ79451.1
DgF3'5'H	<i>Delphinium grandiflorum</i>	AAX51796.1
LeF3'5'H	<i>Lobelia erinus</i>	BAF49321.1
CmF3'5'H	<i>Campanula medium</i>	O04773
EsF3'5'H	<i>Epimedium sagittatum</i>	ADE80942.1
GhiF3'5'H	<i>Gossypium hirsutum</i>	AAP31058
VvF3'5'H	<i>Vitis vinifera</i>	ABH06585
CpF3'5'H	<i>Cyclamen persicum</i>	ACX37698.1
CsF3'5'H	<i>Camellia sinensis</i>	ABA40923.1
CtF3'5'H	<i>Clitoria ternatea</i>	BAE72870
GmF3'5'H	<i>Glycine max</i>	AAM51564.1
VwF3'5'H	<i>Viola wittrockiana</i>	BAF93855.1
EeF3'5'H	<i>Eustoma exaltatum</i>	Q96418
GtF3'5'H	<i>Gentiana triflora</i>	BAA12735
VmF3'5'H	<i>Vinca major</i>	BAC97831.1
CrF3'5'H	<i>Catharanthus roseus</i>	CAA09850
PhF3'5'H (Hf1)	<i>Petunia hybrida</i>	CAA80266.1
PhF3'5'H1 (Hf2)	<i>Petunia hybrida</i>	CAA80265.1
StF3'5'H	<i>Solanum tuberosum</i>	AAV85470.1
SIF3'5'H	<i>Solanum lycopersicum</i>	ADC80513.1
SmeF3'5'H	<i>Solanum melongena</i>	P37120
ThF3'5'H	<i>Torenia hybrida</i>	BAB20076
AkF3'5'H	<i>Antirrhinum kelloggii</i>	BAJ16329.1
GlhF3'5'H	<i>Glandularia hybrida</i>	AAT34974.1

Table S7. The proteins used in the phylogenetic tree of F3H, FLS and ANS

Protein name	Plant Species	Accession number
GbF3H	<i>Ginkgo biloba</i>	AAU93347.1
ZmANS	<i>Zea mays</i>	NP_001106074.1
SbANS	<i>Sorghum bicolor</i>	XP_002451336.1
OsANS	<i>Oryza sativa</i>	CAA69252.1
TaANS	<i>Triticum aestivum</i>	BAE98276.1
AcANS1	<i>Allium cepa</i>	ABR24157.1
AcANS2	<i>Allium cepa</i>	ABM66367.1
SsANS	<i>Solenostemon scutellarioides</i>	ABP57079.1
PfANS	<i>Perilla frutescens</i>	BAA20143.1
AtANS	<i>Arabidopsis thaliana</i>	NP_194019.1
NtANS1	<i>Nicotiana tabacum</i>	AFM52334.1
NtANS2	<i>Nicotiana tabacum</i>	AFM52335.1
PhANS	<i>Petunia hybrida</i>	P51092.1
StANS1	<i>Solanum tuberosum</i>	NP_001274859.1
StANS2	<i>Solanum tuberosum</i>	AEJ90548.1
MtANS	<i>Medicago truncatula</i>	ABU40983.1
GmANS	<i>Glycine max</i>	AAR26525.1
EgANS	<i>Eustoma grandiflorum</i>	BAJ08929.2
TcANS	<i>Theobroma cacao</i>	ADD51355.1
VvANS	<i>Vitis vinifera</i>	NP_001268147.1
CsANS	<i>Citrus sinensis</i>	NP_001275784.1
FaANS	<i>Fragaria ananassa</i>	AAU12369.1
PpANS	<i>Prunus persica</i>	ABX89943.1
MdANS	<i>Malus domestica</i>	AAD26205.1
PcANS	<i>Pyrus communis</i>	ABB70119.1
GmF3H1	<i>Glycine max</i>	NP_001236797.1
GmF3H2	<i>Glycine max</i>	AAU06218.1
VvF3H	<i>Vitis vinifera</i>	CAA53579.1
AtF3H	<i>Arabidopsis thaliana</i>	NP_190692.1
MtF3H	<i>Medicago truncatula</i>	ACR15123.1
StF3H	<i>Solanum tuberosum</i>	NP_001274930.1
PhF3H	<i>Petunia hybrida</i>	AAC49929.1
CsF3H	<i>Citrus sinensis</i>	BAA36553.1
FaF3H	<i>Fragaria ananassa</i>	BAE17126.1
PpF3H	<i>Prunus persica</i>	AQX36284.1
PcF3H	<i>Pyrus communis</i>	AGL81347.1
MdF3H	<i>Malus domestica</i>	BAB92997.1
EgF3H	<i>Eustoma grandiflorum</i>	BAD34459
PcrF3H	<i>Petroselinum crispum</i>	AAP57394.1
AcF3H	<i>Allium cepa</i>	AAO63022.1
NtF3H	<i>Nicotiana tabacum</i>	AAC15414.1

Table S7. Cont.

AmF3H	<i>Antirrhinum majus</i>	BAX37181.1
SsF3H	<i>Solenostemon scutellarioides</i>	ABP57073.1
PtF3H	<i>Perilla frutescens</i>	BAA19657.1
OsF3H	<i>Oryza sativa</i>	NP_001054157.1
TaF3H	<i>Triticum aestivum</i>	ABR13013.1
SbF3H	<i>Sorghum bicolor</i>	ADB66755.1
ZmF3H1	<i>Zea mays</i>	NP_001130275.1
ZmF3H2	<i>Zea mays</i>	AAA91227.1
TcF3H	<i>Theobroma cacao</i>	EOX90852.1
GbF3H	<i>Ginkgo biloba</i>	AAU93347.1
MdFLS	<i>Malus domestica</i>	AAX89401.1
PpFLS	<i>Prunus persica</i>	AJO70134.1
PcFLS	<i>Pyrus communis</i>	ABB70118.1
FaFLS	<i>Fragaria ananassa</i>	ABH07784.1
MtFLS1	<i>Medicago truncatula</i>	AES97739.1
MtFLS2	<i>Medicago truncatula</i>	AES71332.1
GmFLS	<i>Glycine max</i>	NP_001237419.1
VvFLS	<i>Vitis vinifera</i>	BAE75807.1
AmFLS	<i>Antirrhinum majus</i>	ABB53382.1
EgFLS	<i>Eustoma grandiflorum</i>	BAD34463.1
PcrFLS	<i>Petroselinum crispum</i>	AAP57395.1
StFLS	<i>Solanum tuberosum</i>	CAA63092.1
PhFLS	<i>Petunia hybrida</i>	CAA80264.1
NtFLS	<i>Nicotiana tabacum</i>	ABE28017.1
AcFLS	<i>Allium cepa</i>	AQR58516.1
TcFLS	<i>Theobroma cacao</i>	EOY09743.1
CsFLS1	<i>Citrus sinensis</i>	XP_006485472.1
CsFLS2	<i>Citrus sinensis</i>	XP_006466183.1
AtFLS1	<i>Arabidopsis thaliana</i>	NP_196481.1
AtFLS2	<i>Arabidopsis thaliana</i>	NP_201163.1
AtFLS3	<i>Arabidopsis thaliana</i>	NP_201164.1
AtFLS4	<i>Arabidopsis thaliana</i>	NP_680463.1
AtFLS5	<i>Arabidopsis thaliana</i>	NP_001032131.1
AtFLS6	<i>Arabidopsis thaliana</i>	NP_680388.1
OsFLS	<i>Oryza sativa</i>	BAD17324.1
TaFLS	<i>Triticum aestivum</i>	AHW49495.1
ZmFLS	<i>Zea mays</i>	NP_001140915.1
SbFLS	<i>Sorghum bicolor</i>	EES07584.1
GbFLS	<i>Ginkgo biloba</i>	ACY00393.1

Table S8. The proteins used in the phylogenetic tree of DFR

Protein name	Plant Species	Accession number
FhDFR1	<i>Freesia hybrida</i>	APG32498.1
FhDFR2	<i>Freesia hybrida</i>	APG32494.1
FhDFR3	<i>Freesia hybrida</i>	APG32495.1
IhDFR	<i>Iris hollandica</i>	BAF93856.1
ApDFR	<i>Agapanthus praecox</i>	BAE78769.1
BfDFR	<i>Bromheadia finlaysoniana</i>	AAB62873.1
LhDFR	<i>Lilium hybrida</i>	BAB40789.1
ZmDFR	<i>Zea mays</i>	NP_001152467.2
OsDFR	<i>Oryza sativa</i>	BAA36182.1
HvDFR	<i>Hordeum vulgare</i>	P51106.1
TaDFR1	<i>Triticum aestivum</i>	AAO60213.1
TaDFR2	<i>Triticum aestivum</i>	AAQ77347.1
AtrDFR	<i>Amborella trichopoda</i>	XP_006878526.1
GbDFR	<i>Ginkgo biloba</i>	AGR34043.1
LjDFR1	<i>Lotus japonicus</i>	BAE19948.1
LjDFR2	<i>Lotus japonicus</i>	BAE19949.1
LjDFR3	<i>Lotus japonicus</i>	BAE19950.1
LjDFR4a	<i>Lotus japonicus</i>	BAE19951.1
LjDFR4b	<i>Lotus japonicus</i>	BAE19952.1
LjDFR5	<i>Lotus japonicus</i>	BAE19953.1
MtDFR1	<i>Medicago truncatula</i>	AAR27014.1
MtDFR2	<i>Medicago truncatula</i>	AAR27015.1
GmDFR1	<i>Glycine max</i>	NP_001238612.2
GmDFR2	<i>Glycine max</i>	ABM64803.1
RhDFR	<i>Rosa hybrida</i>	BAA12723.1
FaDFR	<i>Fragaria ananassa</i>	AAC25960.1
MdDFR	<i>Malus domestica</i>	AAO39816.1
VvDFR	<i>Vitis vinifera</i>	NP_001268144.1
AtDFR	<i>Arabidopsis thaliana</i>	BAA85261.1
DcDFR	<i>Dianthus caryophyllus</i>	CAA91924.1
GhDFR1	<i>Gossypium hirsutum</i>	AHG97389.1
GhDFR2	<i>Gossypium hirsutum</i>	ACV72642.1
PtDFR1	<i>Populus trichocarpa</i>	XP_002300759.1
PtDFR2	<i>Populus trichocarpa</i>	XP_002307667.2
PhDFR	<i>Petunia hybrida</i>	AAF60298.1
NtDFR	<i>Nicotiana tabacum</i>	ABN80437.1
SlDFR	<i>Solanum lycopersicum</i>	NP_001234408.1
IpDFR	<i>Ipomoea nil</i>	BAA22072.1
IpDFR	<i>Ipomoea purpurea</i>	BAA74700.1
CsDFR	<i>Camellia sinensis</i>	BAA84940.1
VmDFR	<i>Vaccinium macrocarpon</i>	AAL89714.1

Table S8. Cont.

GjDFR	<i>Gerbera jamesonii</i>	AHF58605.1
GhyDFR	<i>Gerbera hybrid cultivar</i>	CAA78930.1
AmDFR	<i>Antirrhinum majus</i>	CAA33543.1
AaDFR	<i>Angelonia angustifolia</i>	AHM27144.1
PbDFR	<i>Penstemon barbatus</i>	AIY51701.1
PfDFR	<i>Perilla frutescens</i>	BAA19658.1
SsDFR	<i>Solenostemon scutellarioides</i>	ABP57077.1

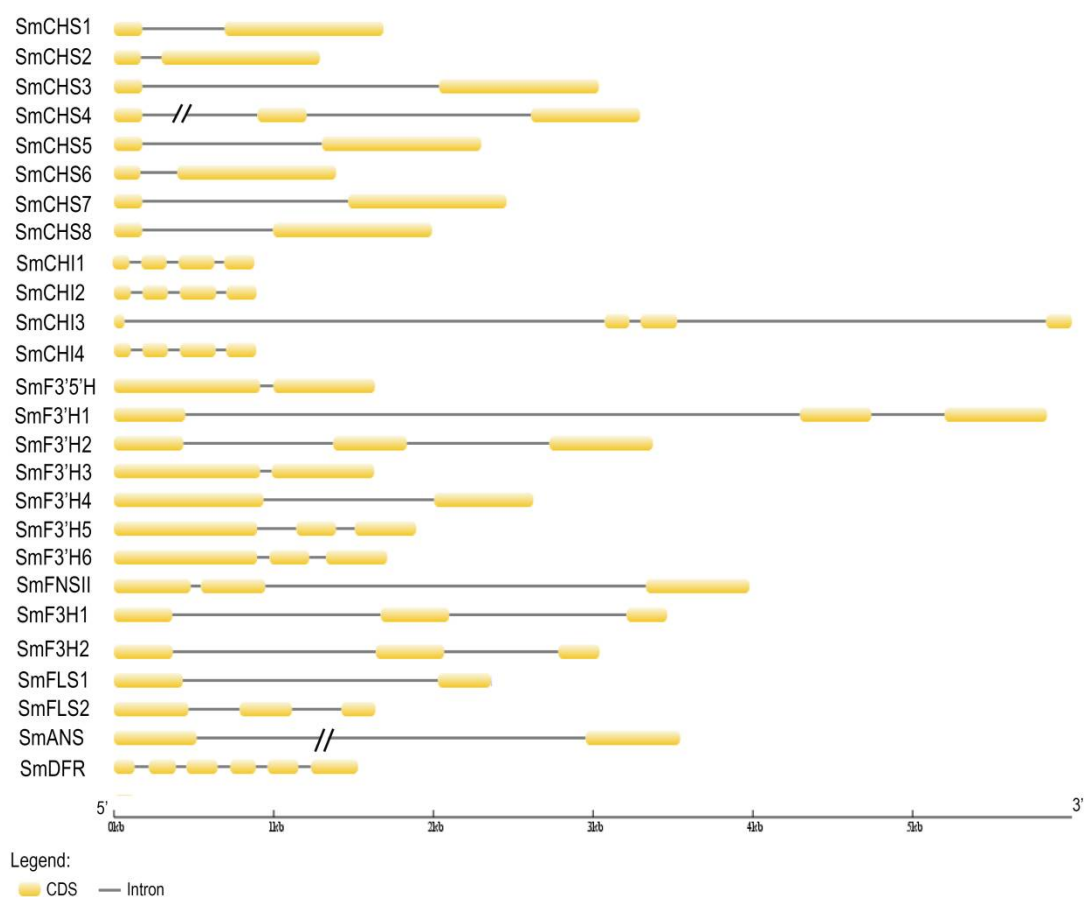


Figure S1. Exon/intron Structures of flavonoid biosynthesis related genes. Thick yellow lines represent exons. Thin gray lines represent introns. Unknown sequences in introns of *SmCHS4* and *SmANS* are indicated by '//'.

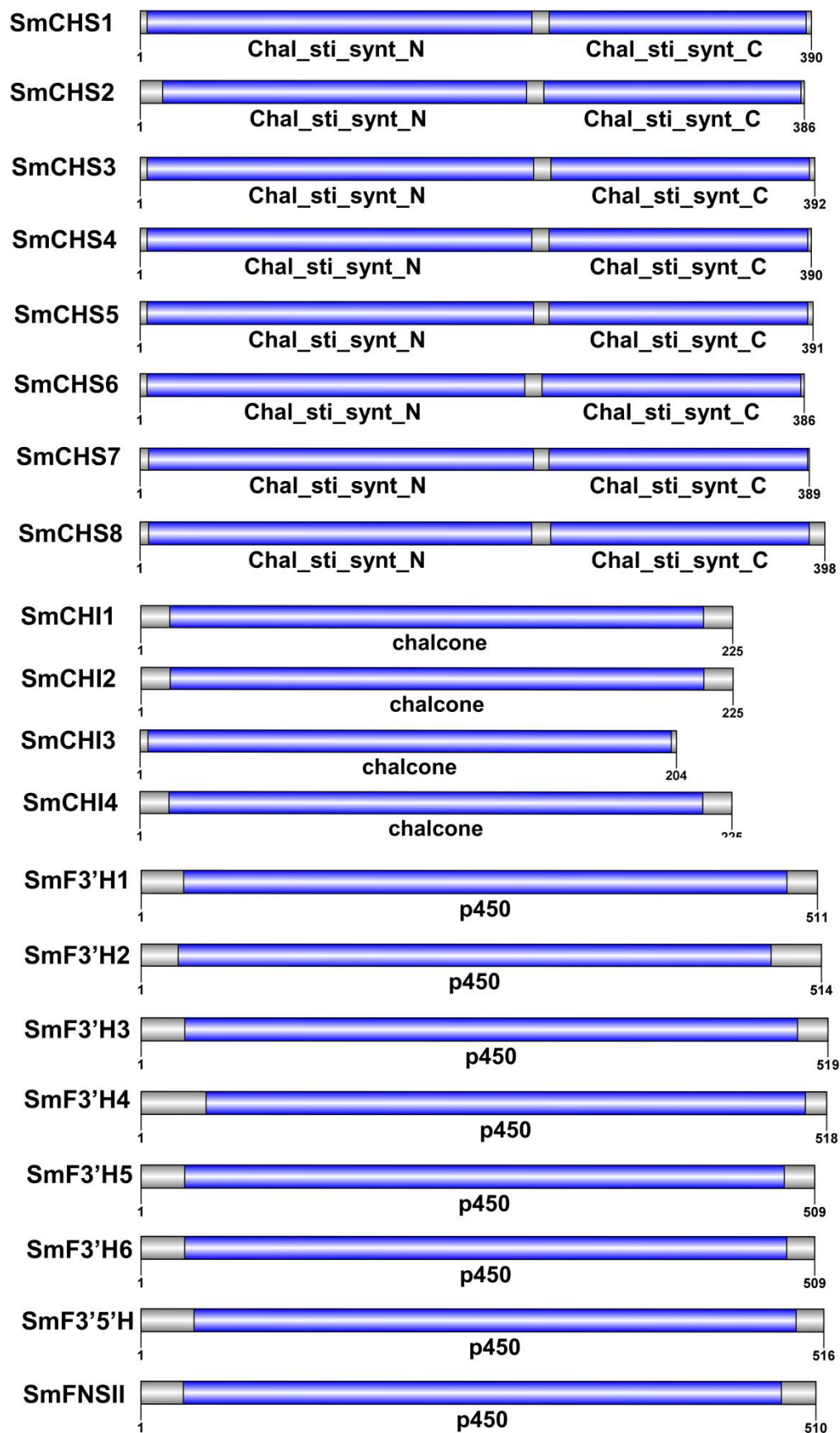


Figure S2. *Cont.*

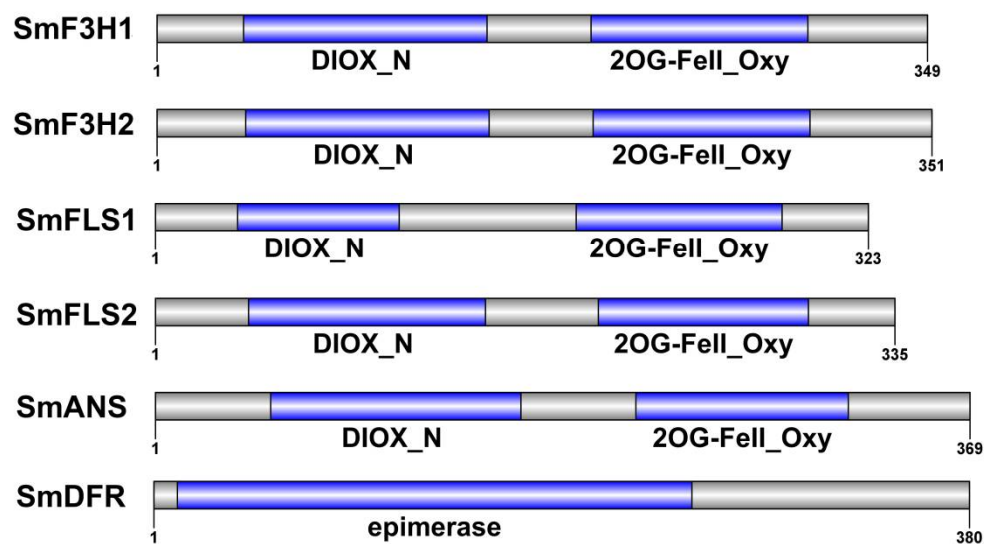


Figure S2. Conserved domains in flavonoid biosynthesis related proteins in *S. miltiorrhiza*. Conserved domains were predicated by searching Pfam and shown in blue boxes. Names of conserved domains are indicated below the boxes.

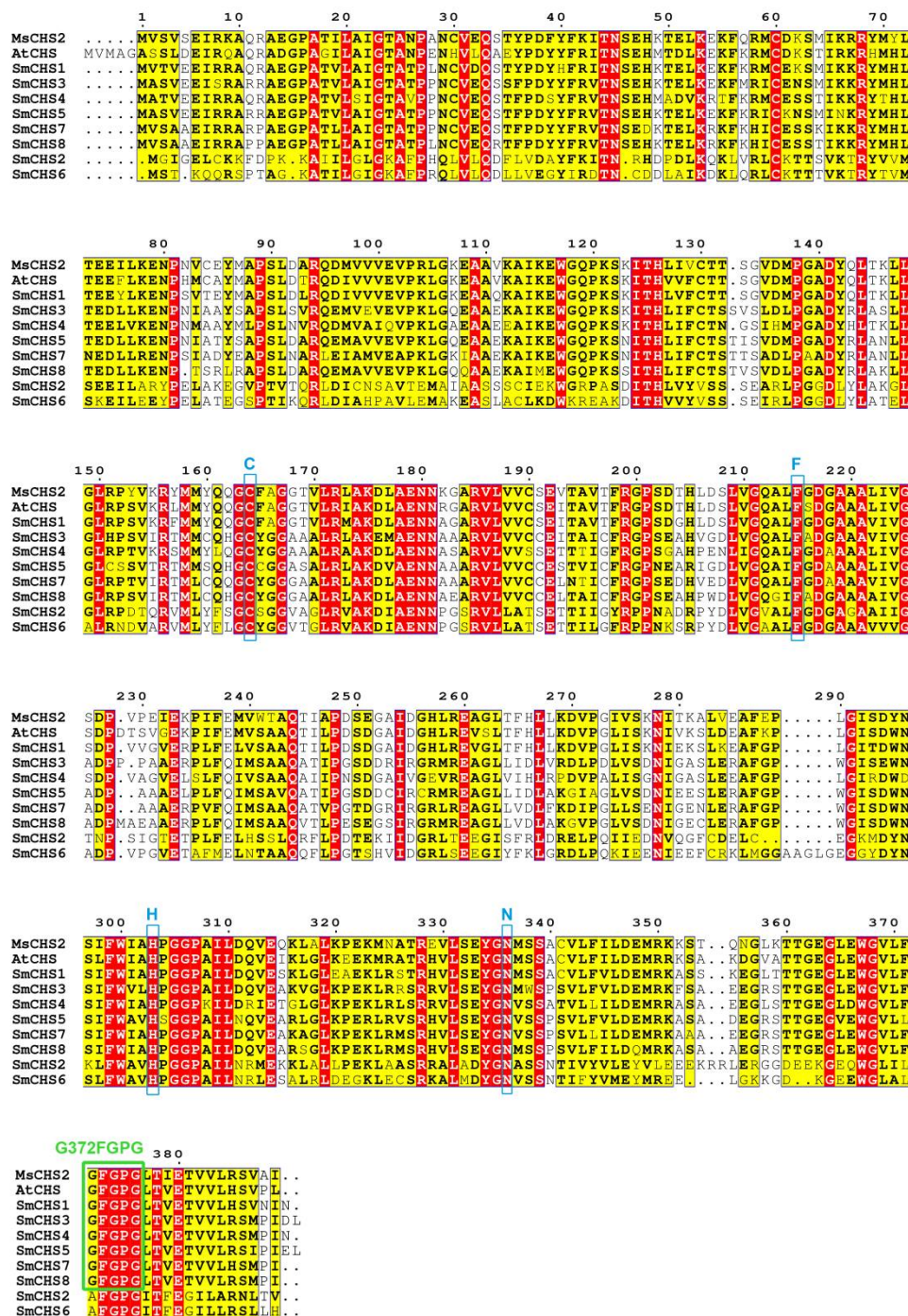


Figure S3. Amino acid sequence alignment of SmCHS1 –SmCHS8 against *Medicago sativa* MsCHS2, and *Arabidopsis thaliana* AtCHS. Identical sites were shown in white letters on a red background, conserved sites were shown in black letters on a yellow background and others were shown in white background. The catalytic triad C164-H303-N336 and the gatekeeper F215 were boxed and marked in blue letters. The highly conserved CHS signature sequence, G372FGPG were boxed and marked in green letters. The figure was produced using the ESPript server (<http://esprict.ibcp.fr/ESPript/ESPript/>).

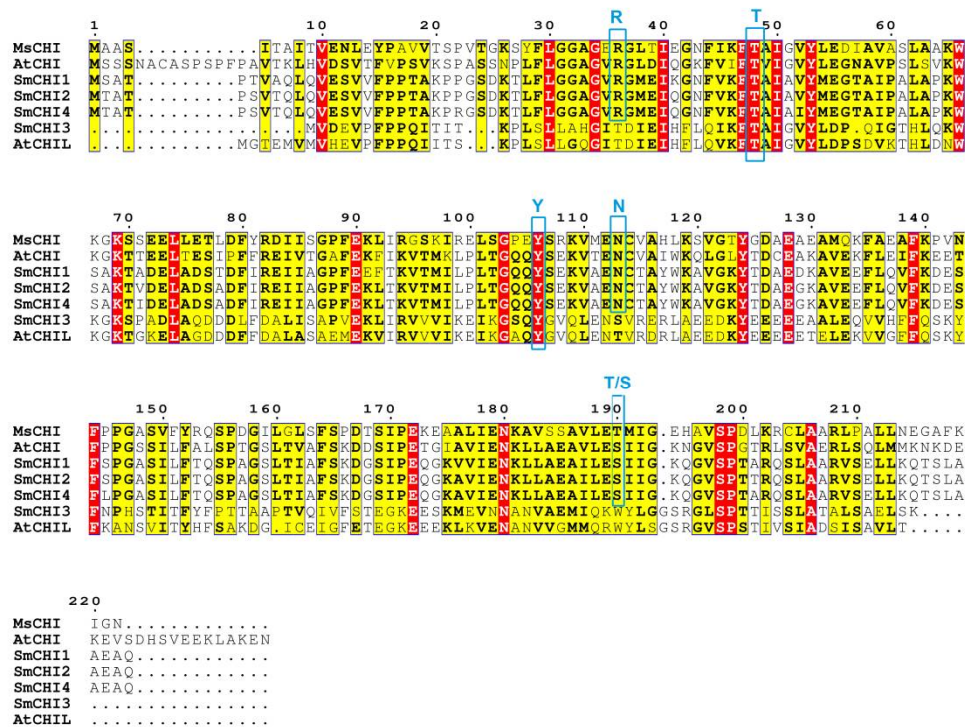


Figure S4. Amino acid sequence alignment of SmCHI1–SmCHI4 against *Medicago sativa* MsCHI, *Arabidopsis thaliana* AtCHI and AtCHIL. Residue numbering refers to the MsCHI sequence. Identical sites were shown in white letters on a red background, conserved sites are shown in black letters on a yellow background and others are shown in white background. Critical catalytic residue of *bona fide* CHI are boxed and indicated in blue letters. Among them, T190 are legume specific whereas S190 are much more conserved in other non-legume plant species. The figure was produced using the ESPrpt server (<http://esprpt.ibcp.fr/ESPrpt/ESPrpt/>)

Proline-rich hinge region

	1	10	20	30	40	50	60
GhyFNSII	...MNTLQLI	FLFF	FPTL	LFVCLPYKRN	NHRRLPSP	PPSP	FPIIGHLHLLGPLI
GmFNSII	...MISESLVVL	LIVFISASL	LKLLFVRENKPK	AHLKNPPSP	PAIP	IPITIGHLHLLKPLI	HHSEFRLDSIRYGPIL
LjFNSII-1.1	MWIFDL	TISFTTLLFL	ITTTALLLL	LKVFKKNNKLR	.PPPS	PFTLPITIGHLHLLGPLI	HQSEHRLSTLYGPIL
LjFNSII-2.1	MWIFDL	TISFTTLLFL	ITTTALLLL	LKVFKKNNKLR	.PPPS	PFTLPITIGHLHLLGPLI	HQSEHRLSTLYGPIL
LmFNSII-1.1	MLIFDL	TISFTTLLFL	ITTTALLLL	LKVFKKNNKLR	.PPPS	PFTLPITIGHLHLLGPLI	HQSEHRLSTLYGPIL
AmFNSII	...MSTLVYS	TLFILSTLL	LTLLTRTRKTRPP	G	PLALPL	ICHLHLLGPKLHH	THQFSQRYGPIL
SbaFNSII-1	...MDLVE	VTLYAALFLL	SAAFLLLIFAGD	.RSSPP	G	PFFLPITIGHLHLLGPKL	HQSEHGLSQRHGCLM
SbaFNSII-2	...MEVTLN	VALLLSAAV	CLMVFTGKRRRR	LPN...PPG	PFFLP	LIGNLNLVS	PRLHHTHMLAQRYGPIM
SmFNSII	...MELVEM	GAYALFLV	SAAFSRSL	LRSKLRRHAP	G	PFFLPITIGHLHLLGPKL	HQSEHLSQRYGPIM

	70	80	90	100	110	120	130	140	
GhyFNSII	HERLGS	VP	CVVST	PD	LAKDF	LKTNEL	AFSSSRKHS	LATDHIT	YG.VAF
GmFNSII	SLRIGS	VK	FIVAST	PS	LAKDF	LKTNEL	TYSSRKNNMAI	NMVTY	HNA
LjFNSII-1.1	QLKIGY	IP	CVVAST	PE	LAKDF	LKTNEL	AFSSSRKHS	AAIKLL	TYD.V
LjFNSII-2.1	QLKIGY	IP	CVVAST	PE	LAKDF	LKTNEL	AFSSSRKHS	AAIKLL	TYD.V
LmFNSII-1.1	QLKIGY	IP	CVVAST	PE	LAKDF	LKTNEL	AFSSSRKHS	AAIKLL	TYD.V
AmFNSII	QLYLG	SV	CVVAST	PE	LAKDF	LKTNEL	AFSSSRKHS	TAIDIV	TYD.S
SbaFNSII-1	QIRLGS	IN	CVVAST	PE	LAKDF	LKTNEL	AFSSSRKHS	TAIDIV	TYN.S
SbaFNSII-2	KFRLGS	IP	CVVST	PD	LAKDF	LKTNEL	AFSSSRKHS	TAIDIV	TYG.V
SmFNSII	QIRLGS	IN	CVVAST	PE	LAKDF	LKTNEL	AFSSSRKHS	TAIDIV	TYD.S

	150	160	170	180	190	200	210
GhyFNSII	LSHF	LP	IR	TH	EL	IR	LMVKS
GmFNSII	LGHF	LP	IR	TH	EL	IR	LMVKS
LjFNSII-1.1	MNHFL	LP	IR	TH	EL	IR	LMVKS
LjFNSII-2.1	MNHFL	LP	IR	TH	EL	IR	LMVKS
LmFNSII-1.1	MNHFL	LP	IR	TH	EL	IR	LMVKS
AmFNSII	LSHF	LP	IR	TH	EL	IR	LMVKS
SbaFNSII-1	LHGF	LP	IR	TH	EL	IR	LMVKS
SbaFNSII-2	LHGF	LP	IR	TH	EL	IR	LMVKS
SmFNSII	LHGF	LP	IR	TH	EL	IR	LMVKS

	220	230	240	250	260	270	280
GhyFNSII	GDFN	VSDF	IWFCK	NID	QGF	KRYE	GT
GmFNSII	GDFN	VSDF	IWFCK	NID	QGF	KRYE	GT
LjFNSII-1.1	GDFN	VSDF	IWFCK	NID	QGF	KRYE	GT
LjFNSII-2.1	GDFN	VSDF	IWFCK	NID	QGF	KRYE	GT
LmFNSII-1.1	GDFN	VSDF	IWFCK	NID	QGF	KRYE	GT
AmFNSII	GDFN	VSDF	IWFCK	NID	QGF	KRYE	GT
SbaFNSII-1	GDFN	VSDF	IWFCK	NID	QGF	KRYE	GT
SbaFNSII-2	GDFN	VSDF	IWFCK	NID	QGF	KRYE	GT
SmFNSII	GDFN	VSDF	IWFCK	NID	QGF	KRYE	GT

Oxygen binding pocket

	290	300	310	320	330	340	350
GhyFNSII	EDGKA	ET	TRDH	IKAL	ILDF	LTAG	TDTTA
GmFNSII	EDGKA	ET	TRDH	IKAL	ILDF	LTAG	TDTTA
LjFNSII-1.1	EDGKA	ET	TRDH	IKAL	ILDF	LTAG	TDTTA
LjFNSII-2.1	EDGKA	ET	TRDH	IKAL	ILDF	LTAG	TDTTA
LmFNSII-1.1	EDGKA	ET	TRDH	IKAL	ILDF	LTAG	TDTTA
AmFNSII	ESGK	AE	TRDH	IKAL	ILDF	LTAG	TDTTA
SbaFNSII-1	ESGK	AE	TRDH	IKAL	ILDF	LTAG	TDTTA
SbaFNSII-2	ESGK	AE	TRDH	IKAL	ILDF	LTAG	TDTTA
SmFNSII	ESGK	AE	TRDH	IKAL	ILDF	LTAG	TDTTA

ExxR motif

	360	370	380	390	400	410	420
GhyFNSII	YIQAI	IKET	FR	LHPP	IP	MLIR	KSTENV
GmFNSII	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
LjFNSII-1.1	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
LjFNSII-2.1	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
LmFNSII-1.1	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
AmFNSII	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
SbaFNSII-1	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
SbaFNSII-2	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
SmFNSII	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV

PERF motif

	360	370	380	390	400	410	420
GhyFNSII	YIQAI	IKET	FR	LHPP	IP	MLIR	KSTENV
GmFNSII	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
LjFNSII-1.1	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
LjFNSII-2.1	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
LmFNSII-1.1	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
AmFNSII	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
SbaFNSII-1	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
SbaFNSII-2	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV
SmFNSII	YTHAI	IKET	FR	LHPP	IP	MLIR	KSTENV

Figure S5. Cont.

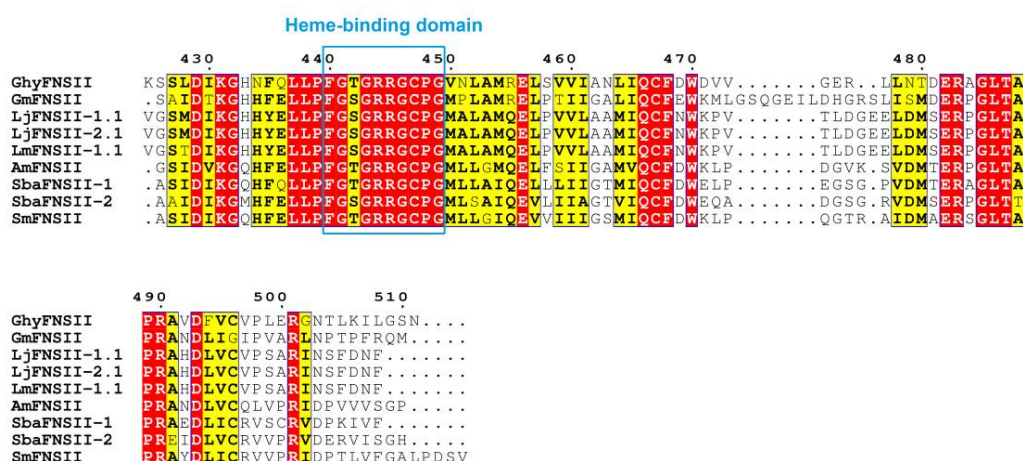


Figure S5. Comparison of SmFNSII proteins with other plant FNSII proteins. The amino acid sequences of SmFNSII was aligned with the amino acid sequences from *Gerbera hybrid* GhyFNSII, *Glycine max* GmFNSII, *Lonicera japonica* LjFNSII-1.1 and LjFNSII-2.1, *Lonicera macranthoides* LmFNSII-1.1, *Antirrhinum majus* AmFNSII and *Scutellaria baicalensis* SbaFNSII-1, SbaFNSII-2. Identical sites were shown in white letters on a red background, conserved sites are shown in black letters on a yellow background and others are shown in white background. P450s have conserved motifs including proline rich hinge region, oxygen binding pocket, E-R-R triade consisting of the E and R from the ExxR in the K-helix and the R from the “PERF” consensus sequence, and heme-binding pocket are boxed in blue. They are universal motifs among P450s. Among them, the E-R-R triad and the cysteine in the heme-binding domain are highly conserved in all primary amino acid sequence of plant P450s. The figure was produced using the ESPript server (<http://esprict.ibcp.fr/ESPript/ESPript/>).

Proline-rich hinge region

	1	10	20	30	40	50	60
AtF3H	M	ATLFL	ATLFL	LILRIFSHRR	NRSHNNR	PPGPNPW	IIGNLPHMGT
PhF3H	MEIL	SLIYTV	IFSLQ	ILRSFFR	KRYPL	PPGPKPW	IIGNLVH
SmF3H1	MQMS	SLTLVCS	AVLGL	LLYSILV	RRPEK	PPGPRPW	IVGNLPOLG
SmF3H2	MSAT	SLIIC	ISILGL	LFYLF	FQKRR	PPGPRPW	IVGNLPOLG
SmF3H3	MQSF	HEFL	GLLIAL	ALLSR	VRIPK	PPGPKPW	IIGNLNL
SmF3H4	MSFL	SNYWSR	WFNTD	DEARL	VFSII	IGLVAV	AWILALLV
SmF3H5	MENQ	ILPIS	LVLT	ITFL	SLHIL	KKILQN	KKLNP
SmF3H6	MENQ	ILPIS	LVLT	ITFL	SLHIL	KKILQN	KKLNP

VVVAAS

	70	80	90	100	110	120	130
AtF3H	TYGPI	LHRL	GFVD	VVVAAS	KSVAE	QFLKI	HANFAS
PhF3H	TYGPI	LMYK	MGFV	VVVAAS	SASVA	QFLKI	HANFAS
SmF3H1	QYVGP	LMHKL	MGFV	VVVAAS	AAVAQ	QFLKA	HANFAS
SmF3H2	RVHGP	LLHKL	MGFV	VVVAAS	AGVAE	QFLRV	HANFAS
SmF3H3	QKYGI	IMLKL	KFGK	PVLV	ASSPH	MAQFL	KVHDA
SmF3H4	KSYGI	PIYTK	LGGKI	GVVIT	SPELAK	QVLKD	DTTFAN
SmF3H5	QKYGI	IMLKL	KFGK	PVLV	ASSPH	MAQFL	KVHDA
SmF3H6	QKYGI	IMLKL	KFGK	PVLV	ASSPH	MAQFL	KVHDA

	140	150	160	170	180	190	200
AtF3H	AKALE	DFKHV	RQEEV	GTITRE	LVVR	GTQPV	NLGGQ
PhF3H	TKALD	DFRHV	RQEEV	GTITRE	LVVR	GTQPV	NLGGQ
SmF3H1	AKALD	DFRHV	RQEEV	GTITRE	LVVR	GTQPV	NLGGQ
SmF3H2	AKALD	DFRHV	RQEEV	GTITRE	LVVR	GTQPV	NLGGQ
SmF3H3	AKRLD	DFKHV	RQEEV	GTITRE	LVVR	GTQPV	NLGGQ
SmF3H4	NATL	DSVYGL	RRLR	QITIGH	HGKAG	TAVD	VGEM
SmF3H5	ARRLS	SEYET	IRAE	EMSS	LLKE	TFGFS	CRPFL
SmF3H6	ARRLS	SEYET	IRAE	EMSS	LLKE	TFGFS	CRPFL

	210	220	230	240	250	260	270
AtF3H	MVTE	EMMAL	AGVNI	GDVFP	SLDW	LDQGV	AKMKR
PhF3H	MVTE	EMMAL	AGVNI	GDVFP	SLDW	LDQGV	AKMKR
SmF3H1	MVTE	EMMAL	AGVNI	GDVFP	SLDW	LDQGV	AKMKR
SmF3H2	MVTE	EMMAL	AGVNI	GDVFP	SLDW	LDQGV	AKMKR
SmF3H3	MLDE	EMMAL	AGVNI	GDVFP	SLDW	LDQGV	AKMKR
SmF3H4	VVGE	MTALL	GAPNI	SDP	PALER	LDQGV	AKMKR
SmF3H5	MLDE	EMMAL	AGVNI	GDVFP	SLDW	LDQGV	AKMKR
SmF3H6	MLDE	EMMAL	AGVNI	GDVFP	SLDW	LDQGV	AKMKR

Oxygen binding pocket motif

	280	290	300	310	320	330	340	350
AtF3H	DLDDGG	SLTDTE	IKALL	NMF	TAAGT	DTSS	TVDW	AIAB
PhF3H	DADNDG	KLTDTE	IKALL	NLF	VAACT	DTSS	TVDW	AIAB
SmF3H1	D.DTQG	KLTDTE	IKALL	NLF	VAACT	DTSS	TVDW	AIAB
SmF3H2	D.NGEG	KLTDTE	IKALL	NLF	VAACT	DTSS	TVDW	AIAB
SmF3H3	P..NLEV	KLTRDC	VKALL	QDL	LTGCT	DTSS	TVDW	AIAB
SmF3H4	H.GDANT	P.TITH	VKALL	MDM	VVGCT	DTSS	TVDW	AIAB
SmF3H5	P..TLEV	KLTRDC	VKALL	QDL	LTGCT	DTSS	TVDW	AIAB
SmF3H6	P..ALEV	KLTRDC	VKALL	QDL	LTGCT	DTSS	TVDW	AIAB

ExxR motif

PERF motif GGEK VDVKG

	360	370	380	390	400	410	420	430
AtF3H	VIKEN	FR	LHP	PT	PSL	PHIA	SESCE	INGY
PhF3H	IVKE	FR	LHP	PT	PSL	PHIA	SESCE	INGY
SmF3H1	VIKEN	FR	LHP	PT	PSL	PHIA	SESCE	INGY
SmF3H2	VIKEN	FR	LHP	PT	PSL	PHIA	SESCE	INGY
SmF3H3	IIMES	SW	RLH	PLA	PL	LAP	HCAME	CKV
SmF3H4	VMKE	AL	RLH	PLA	PL	LAP	HCAME	CKV
SmF3H5	IVKE	FR	LHP	PT	PSL	PHIA	SESCE	INGY
SmF3H6	IVKE	FR	LHP	PT	PSL	PHIA	SESCE	INGY

Heme-binding domain

	440	450	460	470	480	490	500
AtF3H	FELIP	FGAG	GRRIC	AGLS	LGRL	TIOFL	TATLV
PhF3H	FELIP	FGAG	GRRIC	AGLS	LGRL	TIOFL	TATLV
SmF3H1	FELIP	FGAG	GRRIC	AGLS	LGRL	TIOFL	TATLV
SmF3H2	FELIP	FGAG	GRRIC	AGLS	LGRL	TIOFL	TATLV
SmF3H3	FELIP	FGAG	GRRIC	AGLS	LGRL	TIOFL	TATLV
SmF3H4	LNYY	FPFG	GRRIC	AGLS	LGRL	TIOFL	TATLV
SmF3H5	FELIP	FGAG	GRRIC	AGLS	LGRL	TIOFL	TATLV
SmF3H6	FELIP	FGAG	GRRIC	AGLS	LGRL	TIOFL	TATLV

510

AtF3H	GLGSG..
PhF3H	IG....
SmF3H1	HA....
SmF3H2	QAQI..
SmF3H3	QALIVDE
SmF3H4	E.....
SmF3H5	NSM....
SmF3H6	NSM....

Figure S6. Amino acid sequence alignment of SmF3'H1–SmF3'H6 against *Arabidopsis thaliana* AtF3'H from and Petunia hybrid PhF3'H. Identical sites were shown in white letters on a red background, conserved sites are shown in black letters on a yellow background and others are shown in white background. The P450-featured conserved motif, including the proline-rich “hinge” region (P/I)PGPx(G/P)xP, oxygen binding pocket motif, E-R-R triade consisting of the EXXR and PERF motif and heme binding domain are boxed and indicated in blue letters. Three F3'H-specific conserved motifs “VVVAAS”, “GGEK” and “VDVKG” have similarity counterparts at the corresponding sites of SmF3'H1 and SmF3'H2. They are boxed and indicated in green. The figure was produced using the ESPript server (<http://esprpt.ibcp.fr/ESPript/ESPript/>).

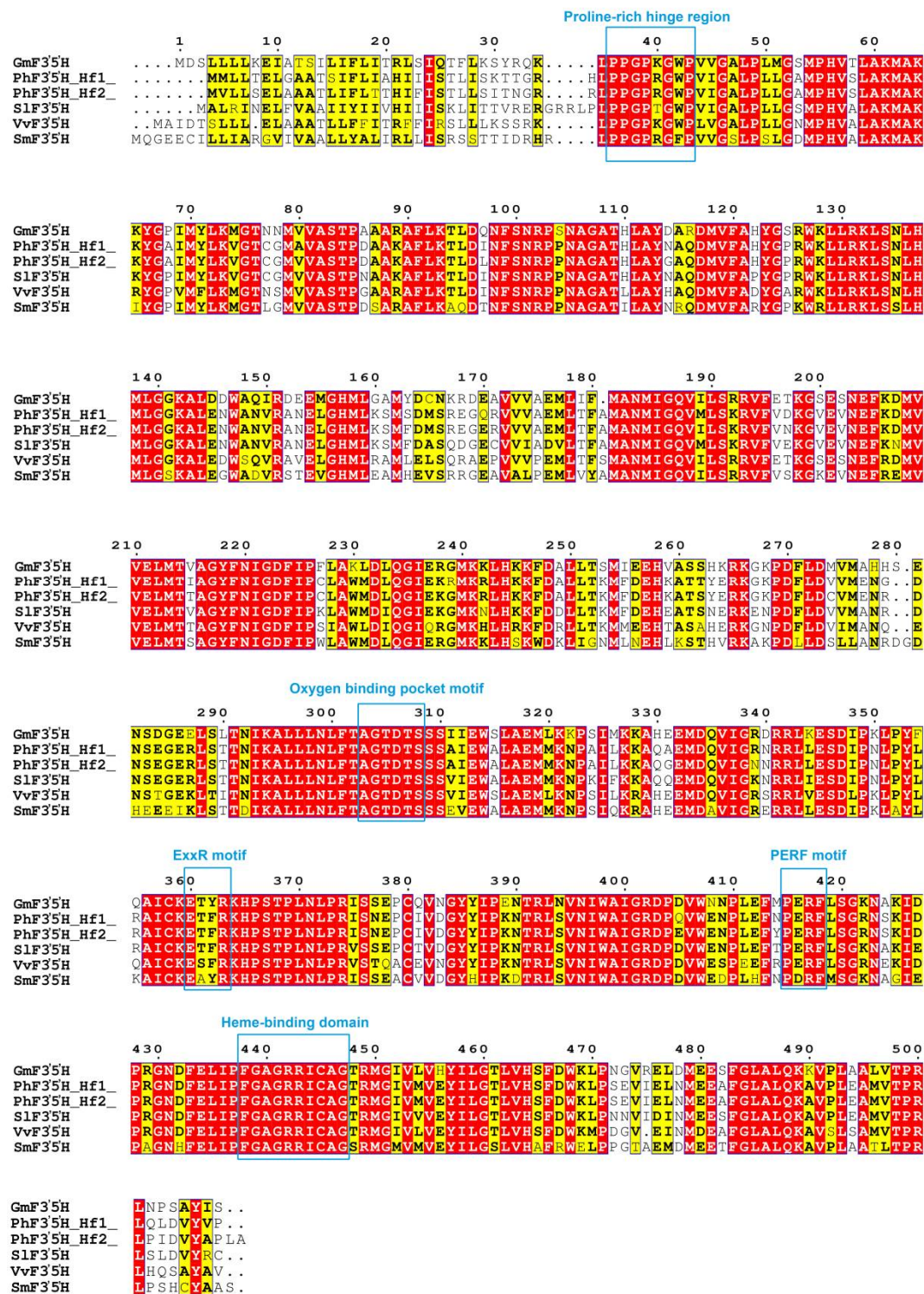


Figure S7. Amino acid sequence alignment of SmF3'5'H against *Glycine max* GmF3'5'H, *Petunia hybrid* PhF3'5'H (Hf1) and PhF3'5'H (Hf2), *Solanum lycopersicum* SlF3'5'H, *Vitis vinifera* VvF3'5'H. The P450-featured conserved motif, including the proline-rich "hinge" region, oxygen binding pocket motif, E-R-R triade consisting of the EXXR and PERF motif and heme binding domain are boxed and indicated in blue letters. The figure was produced using the ESPrpt server (<http://esprpt.ibcp.fr/ESPrpt/ESPrpt/>).

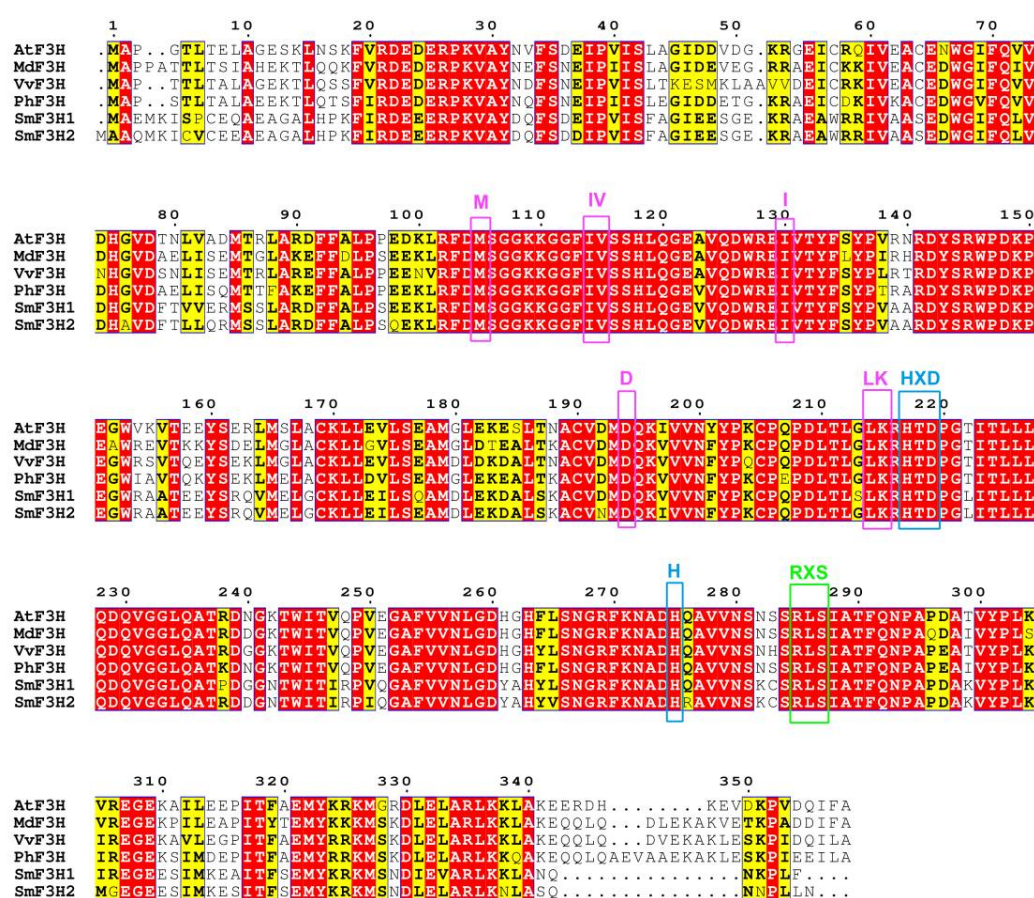


Figure S8. Amino acid sequence alignment of SmF3H1 and SmF3H2 against other plant F3Hs. The amino acid sequences of F3H from *S. miltiorrhiza* was aligned with the amino acid sequences from *Malus domestica* MdF3H, *Vitis vinifera* F3H, and *Petunia hybrid* PhF3H. The amino acid residues involved in ferrus iron binding and 2-oxoglutarate binding are indicated in blue and green box. Seven highly conserved residues with critical roles in determing the activity of F3H are indicated in pink box. The figure was produced using the ESPrpt server (<http://esprpt.ibcp.fr/ESPrpt/ESPrpt/>).

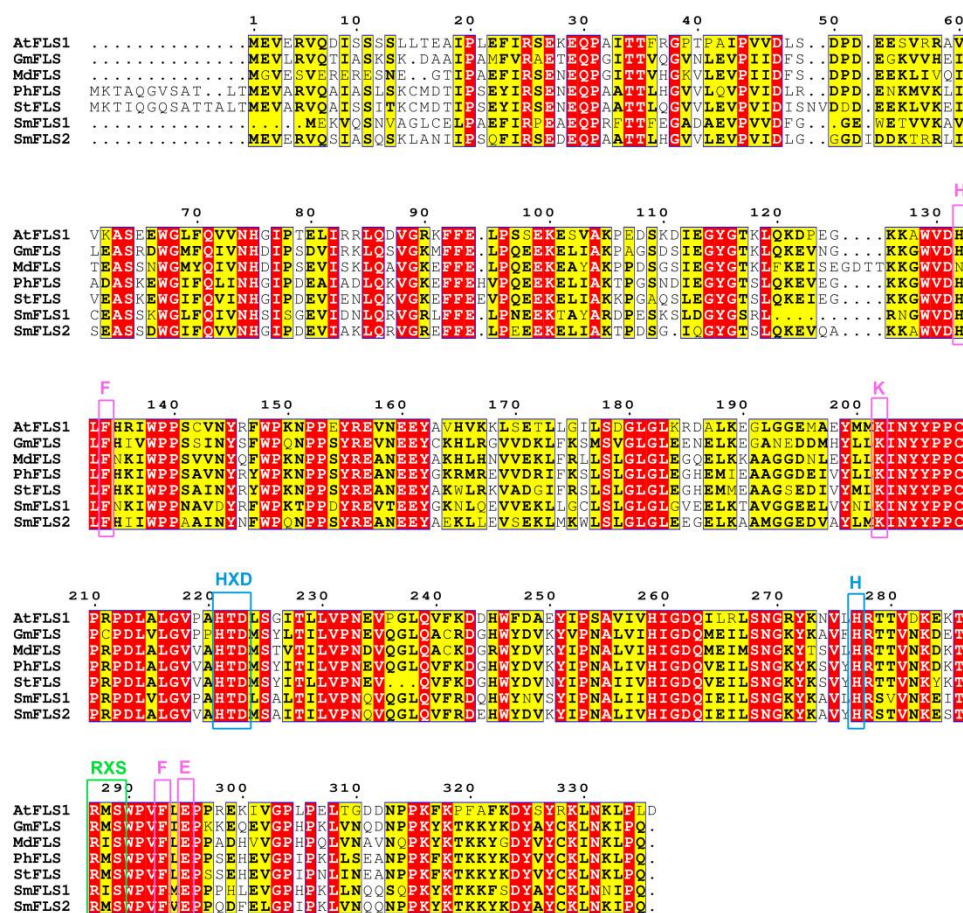


Figure S9. Amino acid sequence alignment of SmFLS1 and SmFLS2 against other plant FLSs. The amino acid sequences of FLS from *S. miltiorrhiza* was aligned with the amino acid sequences from *Malus domestica* MdFLS, *Vitis vinifera* FLS, *Petunia hybrid* PhFLS and *Solanum tuberosum* StFLS. The amino acid residues involved in ferrus iron binding and 2-oxoglutarate binding are indicated in blue and green box, respectively. Five potential substrate binding residues are indicated in pink box. The figure was produced using the ESPript server (<http://esprict.ibcp.fr/ESPript/ESPript/>).

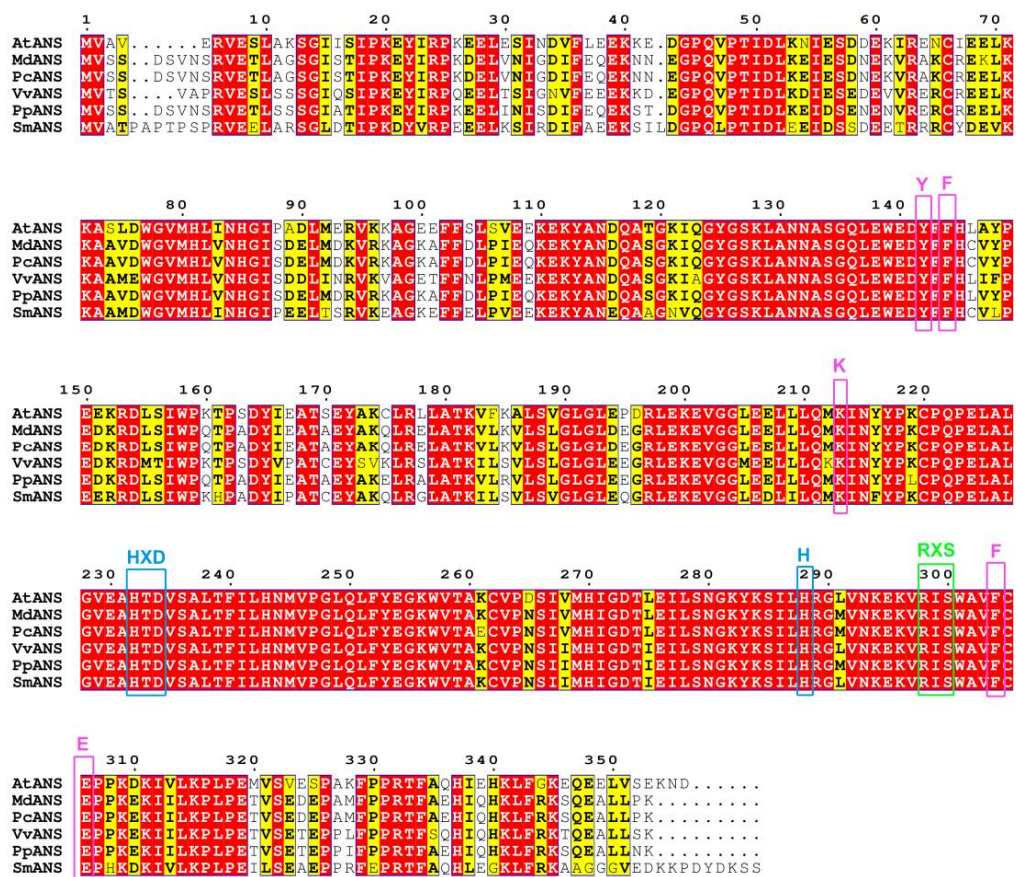


Figure S10. Amino acid sequence alignment of SmANS against other plant ANSs. The amino acid sequences of ANS from *S. miltiorrhiza* was aligned with the amino acid sequences from *Malus domestica* MdANS, *Pyrus communis* PcANS, *Vitis vinifera* VvANS and *Prunus persica* PpANS. The amino acid residues involved in ferrus iron binding and 2-oxoglutarate binding are indicated in blue and green box, respectively. Five potential substrate binding residues are indicated in pink box. The figure was produced using the ESPript server (<http://esprict.ibcp.fr/ESPript/ESPript/>).

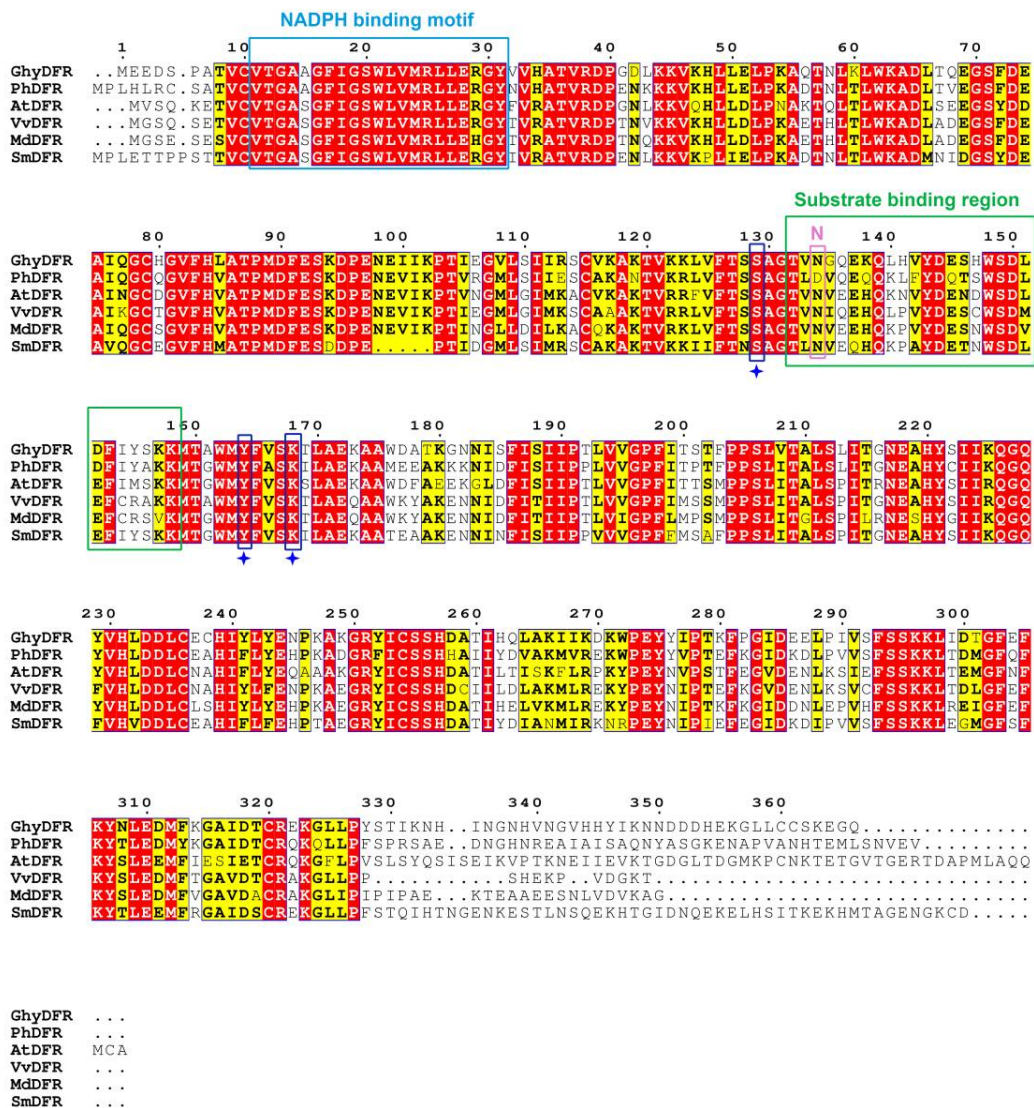


Figure S11. Amino acid sequence alignment of SmDFR against other plant DFRs. The amino acid sequences of DFR from *S. miltiorrhiza* was aligned with the amino acid sequences from *Gerbera hybrid* GhvDFR, *Petunia hybrida* PhDFR, *Arabidopsis thaliana* AtDFR, *Vitis vinifera* VvDFR and *Malus domestica* DFR. The conserved catalytic triad site, S129, Y164 and K168 are boxed in dark blue and indicated in blue stars. NADPH-binding motif is boxed in light blue. The substrate binding region is indicated in green box. The amino acid residue at the 134th position is particularly important for the substrate specificity. It has been indicated in pink box. The figure was produced using the ESPrpt server (<http://esprpt.ibcp.fr/ESPrpt/ESPrpt/>).