

Supplementary Data

Identification of Cytochrome P450 heme motif in Plants Proteome

Akansha Saxena¹, Priyanka Singh², Dharmendra K. Yadav³, Pooja Sharma³, Sarfaraz Alam³, Feroz Khan^{3*}, Sanjog T. Thul⁴, Rakesh K. Shukla⁴, Vikrant Gupta⁴, Neelam S. Sangwan³

Supplementary Table 1. Details of CYP-P450 heme binding sites detected through PPsearch and MEME tools in experimental (known or training) data set.

S.No	Plant name	Protein Sequence ID	Protein's description	Experimental CYP Heme binding site (Prosit ID: PS00086)	Predicted Motif (MEME)
1.	<i>Oryza sativa</i>	gb AAM74394	Putative cytochrome 450	FGAGRRICAG	FGAGRRICAG
2.	<i>Sorghum bicolor</i>	gb ABG54321	Flavonoid3'-hydroxylase	FGAGRRICAG	FGAGRRICAG
3.	<i>Vitis vinifera</i>	dbj BAE47005	Flavonoid3'-hydroxylase	FGAGRRICAG	FGAGRRICAG
4.	<i>Catharanthus roseus</i>	gi 461812	Secologanin synthase, Cytochrome P450	FSWGPRVCLG	FSWGPRVCLG
5.	<i>Arabidopsis thaliana</i>	gi 15225510	Cytochrome P450	FGGRRICPG	FGGRRICPG
6.	<i>Persea americana</i>	gi 117188	Cytochrome P450	FGAGRRGCPG	FGAGRRGCPG
7.	<i>Ageratina adenophora</i>	gb ABM46853	Flavonoid3'-hydroxylase	FGAGRRICVG	FGAGRRICVG
8.	<i>Lobelia erinus</i>	dbj BAF49324	Flavonoid3'-hydroxylase	FGAGRRVCAG	FGAGRRVCAG
9.	<i>Gerbera hybrid cultivar</i>	gb ABA64468	Flavonoid3'-hydroxylase	FGAGRRICVG	FGAGRRICVG
10.	<i>Mentha piperita</i>	emb CAA83941	Cytochrome P450	FGAGRRGCPG	FGAGRRGCPG
11.	<i>Asparagus officinalis</i>	dbj BAD06417	Cytochrome P450	FGAGRRICPG	FGAGRRICPG
12.	<i>Beta vulgaris</i>	gb AAF34537	Isoflavone synthase	FGSGRRMCPG	FGSGRRMCPG
13.	<i>Berberis stolonifera</i>	gb AAC48987	Cytochrome P450	FGSGRRICPG	FGSGRRICPG
14.	<i>Capsicum annum</i>	gb AAF27282	Cytochrome P450	FGAGRRICPG	FGAGRRICPG
15.	<i>Coptis japonica</i>	dbj BAB12433	N-methylcoclaurine-3'-hydroxylase	FGGRRICPG	FGGRRICPG
16.	<i>Gentiana triflora</i>	dbj BAA12735	Flavonoid3',5'-hydroxylase	FGAGRRICAG	FGAGRRICAG
17.	<i>Perilla frutescens var. crispa</i>	dbj BAB59005	Flavonoid3'-hydroxylase	FGSGRRICAG	FGSGRRICAG
18.	<i>Ruta graveolens</i>	emb CAC35977	Putative cinnamate 4-hydroxylase	FGVGRRSCPG	FGVGRRSCPG

Supplementary Table 2. Development of weight matrix of CYP-P450 heme binding motif on the basis of occurrence of each amino acid at that position. Generated consensus sequence represents the frequency of amino acid position.

AA	1	2	3	4	5	6	7	8	9	10
A	0	0	11	0	0	0	0	0	6	0
C	0	0	0	0	0	0	0	18	0	0
D	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0
F	18	0	0	0	0	0	0	0	0	0
G	0	17	2	18	0	0	2	0	0	18
H	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	12	0	0	0
K	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	1	0
M	0	0	0	0	0	0	1	0	0	0
N	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	1	0	0	0	9	0
Q	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	17	18	0	0	0	0
S	0	1	3	0	0	0	1	0	0	0
T	0	0	0	0	0	0	0	0	0	0
V	0	0	1	0	0	0	2	0	2	0
W	0	0	1	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	0
SUM	18	18	18	18	18	18	18	18	18	18
Cons.	F	G	A	G	R	R	I	C	P	G

Supplementary Table 3. Details of highly conserved heme binding site motif predicted through derived weight matrix in the experimental (known) data sets through PoSSuMsearch (validation of training data set).

S. No.	Plant name	Protein ID & Function	Possum Search Results (motif length 10, threshold 61, score range 0-156)					
			Predicted motif	Start position	Score	MSSP (%)	MSS	P-Value
1.	<i>Oryza sativa</i> (japonica cultivar-group)	gi 21672032 gb AAM74394.1 AC119149_9 Putative cytochrome 450	FGAGRRICAG	461	135	86.54	8.653846e-01	1.106967e-10
2.	<i>Sorghum bicolor</i>	gi 110180155 gb ABG54321.1 flavonoid 3'-hydroxylase	FGAGRRICAG	441	135	86.54	8.653846e-01	1.106967e-10
3.	<i>Vitis vinifera</i>	gi 78183422 dbj BAE47005.1 flavonoid 3'-hydroxylase	FGAGRRICAG	436	135	86.54	8.653846e-01	1.106967e
4.	<i>Catharanthus roseus</i>	gi 461812 sp Q05047.1 C72A1_CATRO Cytochrome P450 72A1 (CYPLXXII) (Secologanin synthase) (SLS)	FSWGPRVCLG	462	78	50	5.000000e-01	5.256556e-05
5.	<i>Arabidopsis thaliana</i>	gi 15225510 ref NP_182079.1 CYP76C4 (cytochrome P450, family 76, subfamily C, polypeptide 4); oxygen binding	FGGGRRICPG	442	129	82.69	8.269231e-01	5.238244e-10
6.	<i>Persea americana</i>	gi 117188 sp P24465.1 C71A1_PERAE Cytochrome P450 71A1 (CYPLXXIA1) (ARP-2)	FGAGRRGCPG	435	128	82.05	8.205128e-01	6.42125e-10
7.	<i>Ageratina adenophora</i>	gi 120972543 gb ABM46853.1 flavonoid 3'-hydroxylase	FGAGRRICVG	436	131	83.97	8.397436e-01	1.771575e-10
8.	<i>Lobelia erinus</i>	gi 133874242 dbj BAF49324.1 flavonoid 3'-hydroxylase	FGAGRRVCAG	442	125	80.13	8.012821e-01	1.481797e-09
9.	<i>Gerbera hybrid cultivar</i>	gi 77176700 gb ABA64468.1 flavonoid 3'-hydroxylase	FGAGRRICVG	436	131	83.97	8.397436e-01	1.771575e
10.	<i>Mentha piperita</i>	gi 493475 emb CAA83941.1 cytochrome P-450 oxidase	FGAGRRGCPG	439	128	82.05	8.205128e-01	6.421025e-10
11.	<i>Asparagus officinalis</i>	gi 40645046 dbj BAD06417.1 cytochrome P450	FGAGRRICPG	428	138	88.46	8.846154e-01	3.564742e-11
12.	<i>Beta vulgaris</i>	gi 6979556 gb AAF34537.1 AF195816_1 isoflavone synthase 1	FGSGRRMCPG	426	119	76..28	7.628205e-01	7.263226e-09

13.	<i>Berberis stolonifera</i>	gi 642386 gb AAC48987.1 cytochrome P-450 CYP80	FGSGRRICPG	421	130	83.33	8.333333e-01	2.828595e-10
14.	<i>Capsicum annuum</i>	gi 6739506 gb AAF27282.1 AF122821_1 cytochrome P450	FGAGRRICPG	435	138	88.46	8.846154e-01	3.564742e-11
15.	<i>Coptis japonica</i>	gi 9971208 dbj BAB12433.1 (S)-N-methylcoclaurine-3'-hydroxylase	FGGGRRICPG	419	129	82.69	8.269231e-01	5.238244e-10
16.	<i>Gentiana triflora</i>	gi 1620009 dbj BAA12735.1 flavonoid 3',5'-hydroxylase	FGAGRRICAG	445	135	86.54	8.653846e-01	1.106967e-10
17.	<i>Perilla frutescens var. crispa</i>	gi 14278925 dbj BAB59005.1 flavonoid 3'-hydroxylase	FGSGRRICAG	448	127	81.41	8.141026e-01,	9.341138e-10
18.	<i>Ruta graveolens</i>	gi 13548653 emb CAC35977.1 putative cinnamate 4-hydroxylase	FGVGRRSCPG	439	117	75.00	7.500000e-01	1.632294e-08

Footnote: Motif sequence similarity percentage (MSSP)

Supplementary Table 4. Details of predicted heme binding motif in *Arabidopsis thaliana*. To avoid the false positives, cutoff score used for proteome-wide search through derived weight matrix was ≥ 129 (82.69%). High score motifs belong to CYPs.

S.No	Protein ID	Description	Position	Score	MSS	P-value	E-value	Pattern	MSSP (%)
1.	C71AP_ARAT H Q9STK8	Cytochrome P450 71A25	423	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PA	88.46
2.	C71AL_ARAT H Q9STL2	Cytochrome P450 71A21	424	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PA	88.46
3.	C71AO_ARAT H Q9STK9	Cytochrome P450 71A24	423	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PA	88.46
4.	Q304B3_ARA TH Q304B3	Uncharacterized protein At3g48290.1	425	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PA	88.46
5.	C71AM_ARAT H Q9STL1	Cytochrome P450 71A22	424	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PA	88.46
6.	Q9SAB7_ARA TH Q9SAB7	F25C20.25 protein (Cytochrome P450 like)	447	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PA	88.46
7.	Q9LQ26_ARA TH Q9LQ26	Putative cytochrome P450.	439	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PG	88.46
8.	C76C2_ARAT H O64637	Cytochrome P450 76C2	443	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PG	88.46
9.	Q570F4_ARAT H Q570F4	Cytochrome P450	10	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PG	88.46
10.	Q9LQ25_ARA TH Q9LQ25	Putative cytochrome P450	442	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PG	88.46
11.	Q9ZUQ6_ARA TH Q9ZUQ6	Putative cytochrome p450	447	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PG	88.46
12.	Q9XIQ2_ARA TH Q9XIQ2	Putative cytochrome P450	446	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PG	88.46
13.	Q9XIQ1_ARA TH Q9XIQ1	Putative cytochrome P450	445	138	8.846154e-01	4.327875e-11	8.375869e-04	FGAGRRIC PG	88.46

14.	Q4PSX7_ARA TH Q4PSX7	Putative cytochrome P450	446	138	8.846154e- 01	4.327875e- 11	8.375869e- 04	FGAGRRIC PG	88.46
15.	Q9XIQ3_ARA TH Q9XIQ3	F13O11.23 protein	446	138	8.846154e- 01	4.327875e- 11	8.375869e- 04	FGAGRRIC PG	88.46
16.	C89A2_ARAT H Q42602;Q9XIQ 6	Cytochrome P450 89A2 (CYPLXXXIX) (ATH 6-1)	441	138	8.846154e- 01	4.327875e- 11	8.375869e- 04	FGAGRRIC PG	88.46
17.	Q9FLK1_ARA TH Q9FLK1	Cytochrome P450-like protein	411	138	8.846154e- 01	4.327875e- 11	8.375869e- 04	FGAGRRIC PG	88.46
18.	C71BM_ARAT H Q9LTM1	Cytochrome P450 71B22	433	138	8.846154e- 01	4.327875e- 11	8.375869e- 04	FGAGRRIC PG	88.46
19.	C71BZ_ARAT H Q9LXM3	Cytochrome P450 71B38	433	138	8.846154e- 01	4.327875e- 11	8.375869e- 04	FGAGRRIC PG	88.46
20.	Q1PEI9_ARAT H Q1PEI9	Cytochrome P450 family protein	409	138	8.846154e- 01,	4.327875e- 11	8.375869e- 04	FGAGRRIC PG	88.46
21.	Q9LEX2_ARA TH Q9LEX2	Cytochrome P450 monooxygenase-like protein	431	138	8.846154e- 01	4.327875e- 11	8.375869e- 04	FGAGRRIC PG	88.46
22.	C76C1_ARAT H O64636; O65783	Cytochrome P450 76C1	442	138	8.846154e- 01	4.327875e- 11	8.375869e- 04	FGAGRRIC PG	88.46
23.	F3PH_ARATH Q9SD85	Flavonoid 3'- monooxygenase (Flavonoid 3'- hydroxylase) (AtF3'H) (Cytochrome P450 75B1)	437	135	8.653846e- 01	1.495432e- 10	2.894156e- 03	FGAGRRIC AG	86.53
24.	C71BA_ARAT H Q9LVD2	Cytochrome P450 71B10	435	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PA	83.33

25.	Q84WB3_ARA TH Q84WB3	Putative cytochrome p450	435	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PA	83.33
26.	C71AQ_ARAT H Q9STK7	Cytochrome P450 71A26	423	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PA	83.33
27.	C71BR_ARAT H Q9SAE1	Cytochrome P450 71B27	437	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PA	83.33
28.	C71BQ_ARAT H Q9LTL0	Cytochrome P450 71B26	432	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
29.	Q9SZU1_ARA TH Q9SZU1	Cytochrome P450 monooxygenase-like protein	427	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
30.	Q8VYA6_ARA TH Q8VYA6	Cytochrome P450 monooxygenase-like protein	427	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
31.	C71BO_ARAT H Q9LTL8;Q0W VF2	Cytochrome P450 71B24	434	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
32.	C71B4_ARAT H O65786;Q9LTL 1	Cytochrome P450 71B4	438	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
33.	C71AE_ARAT H P58045	Cytochrome P450 71A14	432	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
34.	C71AF_ARAT	Cytochrome P450	431	130	8.333333e-	3.878698e-	7.506564e-	FGSGRRIC	83.33

	H P58046	71A15			01	10	03	PG	
35.	C71BC_ARAT H Q9ZU07	Cytochrome P450 71B12	428	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
36.	A4VCM5_AR ATH A4VCM5	At5g25130	428	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
37.	Q681L2_ARAT H Q681L2	Cytochrome P450-like protein	224	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
38.	C71BD_ARAT H P58050;Q0WU U4	Cytochrome P450 71B13	428	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
39.	Q9LTM5_ARA TH Q9LTM5	Cytochrome P450	368	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
40.	C71AC_ARAT H O49340	Cytochrome P450 71A12	431	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
41.	Q8LPP9_ARA TH Q8LPP9	At2g30750/T11J7.14	437	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
42.	C71AD_ARAT H O49342	Cytochrome P450 71A13	431	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
43.	C71AI_ARAT H Q9SAB6	Cytochrome P450 71A18	431	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
44.	Q9ZNR0_ARA	Putative cytochrome	466	130	8.333333e-	3.878698e-	7.506564e-	FGSGRRIC	83.33

	TH Q9ZNR0	P450			01	10	03	PG	
45.	C71BH_ARAT H Q9LTM6	Cytochrome P450 71B17	436	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
46.	C71BG_ARAT H Q9LTM7	Cytochrome P450 71B16	436	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
47.	C71B7_ARAT H Q96514	Cytochrome P450 71B7	438	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
48.	C71BS_ARAT H Q9SAE3	Cytochrome P450 71B28	432	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
49.	Q8LB43_ARA TH Q8LB43	Putative cytochrome P450 monooxygenase	432	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
50.	C71BN_ARAT H Q9LTM0;Q6A WT7	Cytochrome P450 71B23	435	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
51.	C71BT_ARAT H Q9SAE4	Cytochrome P450 71B29	432	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
52.	C71BF_ARAT H Q9LW27;Q541 X4	Cytochrome P450 71B15	432	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33
53.	C71B9_ARAT H O64718;Q8VY	Cytochrome P450 71B9	432	130	8.333333e- 01	3.878698e- 10	7.506564e- 03	FGSGRRIC PG	83.33

54.	E9 Q5E922_ARAT H Q5E922	At2g02580	432	130	8.333333e-01	3.878698e-10	7.506564e-03	FGSGRRIC PG	83.33
55.	C71B2_ARAT H O65788;Q4195 3;Q8VZK5;Q9 SAE2	Cytochrome P450 71B2	436	130	8.333333e-01	3.878698e-10	7.506564e-03	FGSGRRIC PG	83.33
56.	C71BJ_ARAT H Q9LTM4	Cytochrome P450 71B19	436	130	8.333333e-01	3.878698e-10	7.506564e-03	FGSGRRIC PG	83.33
57.	Q541W8_ARA TH Q541W8	Putative cytochrome P450	436	130	8.333333e-01	3.878698e-10	7.506564e-03	FGSGRRIC PG	83.33
58.	Q8LCF8_ARA TH Q8LCF8	Cytochrome P450, putative	436	130	8.333333e-01	3.878698e-10	7.506564e-03	FGSGRRIC PG	83.33
59.	C71BK_ARAT H Q9LTM3	Cytochrome P450 71B20	436	130	8.333333e-01	3.878698e-10	7.506564e-03	FGSGRRIC PG	83.33
60.	Q3EB00_ARA TH Q3EB00	Uncharacterized protein At3g26180.2	302	130	8.333333e-01	3.878698e-10	7.506564e-03	FGSGRRIC PG	83.33
61.	Q944I1_ARAT H Q944I1	AT3g26180/MTC11_8	436	130	8.333333e-01	3.878698e-10	7.506564e-03	FGSGRRIC PG	83.33
62.	C71BW_ARAT H Q9LIP5;Q53XE 2	Cytochrome P450 71B35	431	129	8.269231e-01	6.726016e-10	1.301707e-02	FGGRRIC PA	82.69

63.	Q0WVN2_AR ATH Q0WVN2	Putative cytochrome P450	431	129	8.269231e- 01	6.726016e- 10	1.301707e- 02	FGGRRIC PA	82.69
64.	C76C4_ARAT H O64635	Cytochrome P450 76C4	442	129	8.269231e- 01	6.726016e- 10	1.301707e- 02	FGGRRIC PG	82.69
65.	C71BL_ARAT H Q9LTM2	Cytochrome P450 71B21	433	129	8.269231e- 01	6.726016e- 10	1.301707e- 02	FGGRRIC PG	82.69

Footnote: MSSP: motif sequence similarity percentage. E-value: an expected by chance error probability of program to determine the threshold or cutoff for significant motif/hit. E-value calculation is based on P-values, it is simply the P-value times database size (input sequence). P-value: a probability value to determine the threshold or cutoff for significant predicted motif/hit. MSS: a matrix similarity score to determine the threshold or cutoff for predicted motif.

Supplementary Table 5. Details of predicted heme binding motif in *Vitis vinifera*. To avoid the false positives, cutoff score used for proteome-wide search through derived weight matrix was ≥ 135 (86.53%). The putative uncharacterized proteins predicted to be CYP family proteins due to high conservation of heme like motif. High score predicted heme motifs belongs to mostly uncharacterized proteins which indicate their functional role in metabolism, thus belongs to CYP family.

S.No	Protein ID	Description	Position	Score	MSS	P-value	E-value	Pattern	MSSP (%)
1.	A7PB27_VITVI A7PB27_VITVI A7PB27	Chromosome chr16 scaffold_10, whole genome shotgun sequence	44	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPE	88.46
2.	A5ATQ1_VITVI A5ATQ1	Putative uncharacterized protein	455	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPA	88.46
3.	A7NYH0_VITVI A7NYH0	Chromosome chr6 scaffold_3, whole genome shotgun sequence	447	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPA	88.46
4.	A5C4N0_VITVI A5C4N0	Putative uncharacterized protein	422	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPA	88.46
5.	A7PC07_VITVI A7PC07	Chromosome chr2 scaffold_11, whole genome shotgun sequence	433	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46
6.	A5C3R2_VITVI A5C3R2	Putative uncharacterized protein	21	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46
7.	A5AHQ3_VITVI A5AHQ3	Putative uncharacterized protein	434	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46

8.	A5BZH2_VITVI A5BZH2	Putative uncharacterized protein	408	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
9.	A5BZH3_VITVI A5BZH3	Putative uncharacterized protein	377	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
10.	A5AN88_VITVI A5AN88	Putative uncharacterized protein	432	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
11.	A5BXF0_VITVI A5BXF0_VITVI A5BXE0	Putative uncharacterized protein	414	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
12.	A5C9L8_VITVI A5C9L8	Putative uncharacterized protein	56	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
13.	A7PC05_VITVI A7PC05	Chromosome chr2 scaffold_11, whole genome shotgun sequence	164	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
14.	A7PBX3_VITVI A7PBX3	Chromosome chr2 scaffold_11, whole genome shotgun sequence	429	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
15.	A5C7X2_VITVI A5C7X2	Putative uncharacterized protein (Chromosome chr2 scaffold_11, whole genome shotgun sequence).	429	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
16.	A5BV16_VITVI	Putative	350	138	8.846154e-	4.587320e-	6.615024e-	FGAGRRICPG	88.46

	A5BV16	uncharacterized protein			01	11	04		
17.	A7PB24_VITVI A7PB24	Chromosome chr16 scaffold_10, whole genome shotgun sequence	122	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46
18.	A5BAN8_VITVI A5BAN8	Putative uncharacterized protein	31	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46
19.	A7PB17_VITVI A7PB17	Chromosome chr16 scaffold_10, whole genome shotgun sequence	454	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46
20.	A5ARP5_VITVI A5ARP5	Putative uncharacterized protein	36	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46
21.	A7PB12_VITVI A7PB12	Chromosome chr16 scaffold_10, whole genome shotgun sequence	33	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46
22.	A5AS73_VITVI A5AS73	Putative uncharacterized protein	446	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46
23.	A7PAZ8_VITVI A7PAZ8	Chromosome chr16 scaffold_10, whole genome shotgun sequence	446	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46
24.	A7P0F6_VITVI A7P0F6	Chromosome chr19 scaffold_4, whole genome shotgun sequence	219	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46

25.	A7PB00_VITVI A7PB00	Chromosome chr16 scaffold_10, whole genome shotgun sequence	417	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
26.	A5AEC0_VITVI A5AEC0	Putative uncharacterized protein	444	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
27.	A5ADQ1_VITVI A5ADQ1	Putative uncharacterized protein	31	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
28.	A7PB15_VITVI A7PB15	Chromosome chr16 scaffold_10, whole genome shotgun sequence	3	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
29.	A7P486_VITVI A7P486	Chromosome chr1 scaffold_5, whole genome shotgun sequence	438	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
30.	A7P484_VITVI A7P484	Chromosome chr1 scaffold_5, whole genome shotgun sequence	438	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
31.	A7P487_VITVI A7P487	Chromosome chr1 scaffold_5, whole genome shotgun sequence	431	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
32.	A7P483_VITVI A7P483	Chromosome chr1 scaffold_5, whole genome shotgun sequence	368	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46
33.	A5AY94_VITVI	Putative	438	138	8.846154e- 01	4.587320e- 11	6.615024e- 04	FGAGRRICPG	88.46

	A5AY94	uncharacterized protein			01	11	04		
34.	A5BNT5_VITVI A5BNT5	Putative uncharacterized protein	403	138	8.846154e-01	4.587320e-11	6.615024e-04	FGAGRRICPG	88.46
35.	A5AV17_VITVI A5AV17	Putative uncharacterized protein	415	135	8.653846e-01	1.554955e-10	2.242282e-03	FGAGRRICAG	86.53
36.	A5BEX8_VITVI A5BEX8	Putative uncharacterized protein	438	135	8.653846e-01	1.554955e-10	2.242282e-03	FGAGRRICAG	86.53
37.	A5AKY7_VITVI A5AKY7	Putative uncharacterized protein	445	135	8.653846e-01	1.554955e-10	2.242282e-03	FGAGRRICAG	86.53
38.	A5AUQ7_VITVI A5AUQ7	Putative uncharacterized protein	438	135	8.653846e-01	1.554955e-10	2.242282e-03	FGAGRRICAG	86.53
39.	A5BVX3_VITVI A5BVX3	Putative uncharacterized protein	438	135	8.653846e-01	1.554955e-10	2.242282e-03	FGAGRRICAG	86.53
40.	A6XHG1_VITVI A6XHG1	Flavonoid 3'5' hydroxylase	438	135	8.653846e-01	1.554955e-10	2.242282e-03	FGAGRRICAG	86.53
41.	A5BU46_VITVI A5BU46	Putative uncharacterized protein	436	135	8.653846e-01	1.554955e-10	2.242282e-03	FGAGRRICAG	86.53
42.	A5C6P2_VITVI A5C6P2	Putative uncharacterized protein (Chromosome chr17 scaffold_12, whole genome	436	135	8.653846e-01	1.554955e-10	2.242282e-03	FGAGRRICAG	86.53

43.	A5CB42_VITVI A5CB42	shotgun Putative uncharacterized protein	398	135	8.653846e- 01	1.554955e- 10	2.242282e- 03	FGAGRRICAG	86.53
-----	------------------------	---	-----	-----	------------------	------------------	------------------	------------	-------

Supplementary Table 6. Details of predicted heme binding motif in *Zea mays*. To avoid the false positives, cutoff score used for proteome-wide search through derived weight matrix was ≥ 90 (57.69%). The uncharacterized proteins predicted to be CYP family proteins due to high score heme motifs.

S.No	Protein ID	Description	Position	Score	MSS	P-value	E-value	Pattern	MSSP (%)
1.	O04980_MAIZE O04980	Cytochrome P-450	275	135	8.653846e-01	2.283844e-10	3.004895e-04	FGAGRRIC AG	86.53
2.	C71C1_MAIZE Q43250;Q43254	Cytochrome P450 71C1 (Benzoxazineless 4)	462	135	8.653846e-01	2.283844e-10	3.004895e-04	FGAGRRIC AG	86.53
3.	Q8VYA8_MAIZE Q8VYA8	Cytochrome P450 monooxygenase CYP71C3v2	467	130	8.333333e-01	5.475129e-10	7.203726e-04	FGSGRRIC PG	83.33
4.	C71C3_MAIZE P93703;Q43256	Cytochrome P450 71C3 (Benzoxazineless 5)	467	130	8.333333e-01	5.475129e-10	7.203726e-04	FGSGRRIC PG	83.33
5.	C71C2_MAIZE Q43255;O04991	Cytochrome P450 71C2 (Benzoxazineless 3)	472	130	8.333333e-01	5.475129e-10	7.203726e-04	FGSGRRIC PG	83.33
6.	C78A1_MAIZE P48420	Cytochrome P450 78A1 (CYPLXXVIII)	482	128	8.205128e-01	1.152020e-09	1.515734e-03	FGAGRRV CPG	82.05
7.	C71C4_MAIZE Q43257;O04990	Cytochrome P450 71C4	473	119	7.628205e-01	1.210076e-08	1.592119e-02	FGSGRRM CPG	76.28

8.	Q8VYA9_MAIZE Q8VYA9	(Benzoxazineless 2) Cytochrome P450 monooxygenase CYP92A1	443	119	7.628205e- 01	1.210076e- 08	1.592119e-02	FGSGRRM CPG	76.28
9.	Q5GAS0_MAIZE Q5GAS0	Tac7077 (Putative uncharacterized protein tac7077)	67	108	6.923077e- 01	1.879631e- 07	2.473065e-01	LGAGRRF TPS	69.23
10.	A0EVG7_MAIZE A0EVG7	Putative uncharacterized protein tac7077	67	108	6.923077e- 01	1.879631e- 07	2.473065e-01	LGAGRRF TPS	69.23
11.	Q8LGM8_MAIZE Q8LGM8	Cytochrome P450 monooxygenase CYP72A5	465	104	6.666667e- 01	3.677474e- 07	4.838518e-01	FGGGPRIC IG	66.66
12.	Q8VYB0_MAIZE Q8VYB0	Cytochrome P450 monooxygenase CYP72A5	254	104	6.666667e- 01	3.677474e- 07	4.838518e-01	FGGGPRIC IG	66.66
13.	Q8LL74_MAIZE Q8LL74	Cytochrome P450 monooxygenase CYP72A16	468	103	6.602564e- 01	4.483299e- 07	5.898758e-01	FGWGPRIC IG	66.02
14.	Q8LGN4_MAIZE Q8LGN4	Cytochrome P450 monooxygenase CYP72A26	468	103	6.602564e- 01	4.483299e- 07	5.898758e-01	FGWGPRIC IG	66.02
15.	Q7G182_MAIZE Q7G182	Cell wall invertase Incw3	15	100	6.410256e- 01	1.248403e- 06	1.642546e+00	FGAGGRT CIL	64.10
16.	Q9ZTQ5_MAIZE Q9ZTQ5	Cell wall invertase	531	100	6.410256e- 01	1.248403e- 06	1.642546e+00	FGAGGRT CIL	64.10
17.	Q8S421_MAIZE Q8S421	P-type R2R3 Myb protein	126	99	6.346154e- 01	1.593425e- 06	2.096499e+00	PGPGRPGC PS	63.46
18.	C88A1_MAIZE	Cytochrome P450	458	98	6.282051e-	1.743213e-	2.293577e+00	FGLGARL	62.82

	Q43246	88A1 (Dwarf3 protein)			01	06		CPG	
19.	Q94KE5_MAIZE Q94KE5	Cytochrome P450-like protein	482	97	6.217949e-01	1.966699e-06	2.587621e+00	FQAGPRIC LG	62.17
20.	Q9FQA9_MAIZE Q9FQA9	Glutathione S-transferase GST 30	172	94	6.025641e-01	3.346090e-06	4.402511e+00	FGAVGRII GS	60.25
21.	Q7XBD9_MAIZE Q7XBD9	Putative gag-pol	1723	93	5.961538e-01	4.047054e-06	5.324781e+00	FGAPRLIF ES	59.61
22.	Q8VY33_MAIZE Q8VY33	Putative uncharacterized protein tac7077	35	91	5.833333e-01	6.718025e-06	8.839026e+00	AAAGRRF TPS	58.33
23.	Q8W1C5_MAIZE Q8W1C5	Putative uncharacterized protein tac7077	31	91	5.833333e-01	6.718025e-06	8.839026e+00	AAAGRRF TPS	58.33
24.	Q195N6_MAIZE Q195N6	Pyruvate Pi dikinase regulatory protein	19	91	5.833333e-01	6.718025e-06	8.839026e+00	SPAGRRLP PS	58.33
25.	Q8LL72_MAIZE Q8LL72	Cytochrome P450 monooxygenase CYP72A28	369	91	5.833333e-01	6.718025e-06	8.839026e+00	FGWGPRT CIG	58.33
26.	Q5XML0_MAIZE Q5XML0	Proteasome subunit beta type	130	90	5.769231e-01	8.711176e-06	1.146145e+01	FGYGSYFC LS	57.69

Supplementary Table 7. Details of predicted heme binding motif in *Triticum aestivum*. To avoid the false positives, cutoff score used for proteome-wide search through derived weight matrix was ≥ 90 (57.69%).

S.No	Protein ID	Description	Position	Score	MSS	P-value	E-value	Pattern	MSSP (%)
1.	Q5YFS1_WHEAT Q5YFS1	Flavonoid 3'-hydroxylase	61	135	8.653846e-01	1.873889e-10	1.702998e-04	FGAGRRIC AG	86.53
2.	Q58A30_WHEAT Q58A30	Cytochrome P450	460	135	8.653846e-01	1.873889e-10	1.702998e-04	FGAGRRIC AG	86.53
3.	Q58A31_WHEAT Q58A31	Cytochrome P450	460	135	8.653846e-01	1.873889e-10	1.702998e-04	FGAGRRIC AG	86.53
4.	Q5QHT6_WHEAT Q5QHT6	Cytochrome P450	460	135	8.653846e-01	1.873889e-10	1.702998e-04	FGAGRRIC AG	86.53
5.	Q5QHT7_WHEAT Q5QHT7	Cytochrome P450	460	135	8.653846e-01	1.873889e-10	1.702998e-04	FGAGRRIC AG	86.53
6.	Q5QHU0_WHEAT Q5QHU0	Cytochrome P450	460	135	8.653846e-01	1.873889e-10	1.702998e-04	FGAGRRIC AG	86.53
7.	Q5J5B6_WHEAT Q5J5B6	Cytochrome P450	460	135	8.653846e-01	1.873889e-10	1.702998e-04	FGAGRRIC AG	86.53
8.	Q8S9E9_WHEAT Q8S9E9	Cytochrome P450	460	135	8.653846e-01	1.873889e-10	1.702998e-04	FGAGRRIC AG	86.53
9.	Q5QHT8_WHEAT Q5QHT8	Cytochrome P450	460	135	8.653846e-01	1.873889e-10	1.702998e-04	FGAGRRIC AG	86.53
10.	Q675C1_WHEAT Q675C1	Cytochrome P450	460	135	8.653846e-01	1.873889e-10	1.702998e-04	FGAGRRIC AG	86.53
11.	Q58A28_WHEAT Q58A28	Cytochrome P450	456	130	8.333333e-01	4.529083e-10	4.116049e-04	FGSGRRICP G	83.33
12.	Q58A29_WHEAT Q58A29	Cytochrome P450	457	130	8.333333e-01	4.529083e-10	4.116049e-04	FGSGRRICP G	83.33
13.	Q8GZU6_WHEAT Q8GZU6	Cytochrome P450	464	130	8.333333e-01	4.529083e-10	4.116049e-04	FGSGRRICP G	83.33

14.	Q8S9E7_WHEAT Q8S9E7	Cytochrome P450	456	130	8.333333e-01	4.529083e-10	4.116049e-04	FGSGRRICP G	83.33
15.	Q70CT2_WHEAT Q70CT2	Cytochrome P450-like protein	425	128	8.205128e-01	9.783772e-10	8.891531e-04	FGAGRRVC PG	82.05
16.	Q70EV8_WHEAT Q70EV8	Cytochrome P450	430	128	8.205128e-01	9.783772e-10	8.891531e-04	FGAGRRVC PG	82.05
17.	Q70EV9_WHEAT Q70EV9	Cytochrome P450	433	128	8.205128e-01	9.783772e-10	8.891531e-04	FGAGRRVC PG	82.05
18.	Q70EW0_WHEAT Q70EW0	Cytochrome P450	432	128	8.205128e-01	9.783772e-10	8.891531e-04	FGAGRRVC PG	82.05
19.	A7U1X5_WHEAT A7U1X5	Cytochrome P450	449	127	8.141026e-01	1.421970e-09	1.292292e-03	FGTGRRICP G	81.41
20.	Q58A33_WHEAT Q58A33	Cytochrome P450	461	127	8.141026e-01	1.421970e-09	1.292292e-03	FGFGRRICP G	81.41
21.	Q8GZU7_WHEAT Q8GZU7	Cytochrome P450	461	127	8.141026e-01	1.421970e-09	1.292292e-03	FGFGRRICP G	81.41
22.	Q8S9E8_WHEAT Q8S9E8	Cytochrome P450	461	127	8.141026e-01	1.421970e-09	1.292292e-03	FGFGRRICP G	81.41
23.	Q58A32_WHEAT Q58A32	Cytochrome P450	461	127	8.141026e-01	1.421970e-09	1.292292e-03	FGFGRRICP G	81.41
24.	Q9AVM3_WHEAT Q9AVM3	Cytochrome P450	440	126	8.076923e-01	2.007966e-09	1.824848e-03	FGAGRRQC PG	80.7
25.	Q9FVM9_WHEAT Q9FVM9	Cytochrome P450	435	117	7.500000e-01	2.322034e-08	2.110274e-02	FGVGRRSC PG	75.00
26.	Q8S9E6_WHEAT Q8S9E6	Cytochrome P450	463	116	7.435897e-01	2.508613e-08	2.279837e-02	FGFGRRMC PG	74.35
27.	Q58A34_WHEAT Q58A34	Cytochrome P450	463	116	7.435897e-01	2.508613e-08	2.279837e-02	FGFGRRMC PG	74.35
28.	Q8S9E5_WHEAT Q8S9E5	Cytochrome P450	463	116	7.435897e-01	2.508613e-08	2.279837e-02	FGFGRRMC PG	74.35

29.	Q3MV21_WHEAT Q3MV21	Fructan exohydrolase	533	104	6.666667e-01	3.191189e-07	2.900165e-01	FGSGGRICI T	66.66
30.	Q2V064_WHEAT Q2V064	Cytochrome P450	471	103	6.602564e-01	3.932596e-07	3.573959e-01	FGWGPRICI G	66.02
31.	Q2V067_WHEAT Q2V067	Cytochrome P450	471	103	6.602564e-01	3.932596e-07	3.573959e-01	FGWGPRICI G	66.02
32.	Q2V065_WHEAT Q2V065	Cytochrome P450	454	103	6.602564e-01	3.932596e-07	3.573959e-01	FGWGPRICI G	66.02
33.	Q2V066_WHEAT Q2V066	Cytochrome P450	471	103	6.602564e-01	3.932596e-07	3.573959e-01	FGWGPRICI G	66.02
34.	Q2QKB0_WHEAT Q2QKB0	Alternative splicing regulator	94	94	6.025641e-01	2.945101e-06	2.676519e+00	GGGGRRGG VS	62.25
35.	CP51_WHEAT P93596;P93595	Cytochrome P450 51 (CYPLI) (P450-LIA1) (Obtusifoliol 14-alpha demethylase)	387	93	5.961538e-01	3.589387e-06	3.262049e+00	FGGGRHGC LG	59.61
36.	O81119_WHEAT O81119	Cell wall invertase	413	93	5.961538e-01	3.589387e-06	3.262049e+00	FGGGGRVC IT	59.61
37.	Q84LA1_WHEAT Q84LA1	Fructan 1- exohydrolase precursor	536	93	5.961538e-01	3.589387e-06	3.262049e+00	FGGGGRVC IT	59.61
38.	Q84PN8_WHEAT Q84PN8	Fructan 1- exohydrolase w1 precursor	537	93	5.961538e-01	3.589387e-06	3.262049e+00	FGGGGRVC IT	59.61
39.	Q70LR2_WHEAT Q70LR2	Fructan 1- exohydrolase w3 precursor	536	93	5.961538e-01	3.589387e-06	3.262049e+00	FGGGGRVC IT	59.61

40.	Q4W8R0_WHEAT Q4W8R0	Fructan exohydrolase	535	93	5.961538e- 01	3.589387e- 06	3.262049e+00	FGGGGRVC IM	59.61
41.	Q4W8R1_WHEAT Q4W8R1	Fructan exohydrolase		93	5.961538e- 01	3.589387e- 06	3.262049e+00	FGGGGRVC IM	59.61
42.	Q7X9L7_WHEAT Q7X9L7	H-protein	173	92	5.897436e- 01	4.543936e- 06	4.129547e+00	RGAGRRGL PR	58.97
43.	Q00NV2_WHEAT Q00NV2	Resistance- related receptor-like kinase	244	91	5.833333e- 01	6.163772e- 06	5.601661e+00	EGSGRRCA FS	58.33
44.	Q2QKC2_WHEAT Q2QKC2	Pre-mRNA processing factor	110	91	5.833333e- 01	6.163772e- 06	5.601661e+00	LGSGRRRS RS	58.33
45.	Q2QKC3_WHEAT Q2QKC3	Pre-mRNA processing factor	120	91	5.833333e- 01	6.163772e- 06	5.601661e+00	LGSGRRRS RS	58.33
46.	Q2UXF7_WHEAT Q2UXF7	Fructan 6- exohydrolase	542	91	5.833333e- 01	6.163772e- 06	5.601661e+00	FGGGGRTCI T	58.33
47.	Q5XUV7_WHEAT Q5XUV7	Proteasome subunit beta type	130	90	5.769231e- 01	7.966289e- 06	7.239795e+00	FGYGSYFC LS	57.69
48.	A1YJD9_WHEAT A1YJD9	WRKY protein	138	90	5.769231e- 01	7.966289e- 06	7.239795e+00	RGGGRRRT GS	57.69

Supplementary Table 8. Details of predicted heme binding motif in *Sorghum bicolor*. To avoid the false positives, cutoff score used for proteome-wide search through derived weight matrix was ≥ 135 (86.53%). All predicted motifs found in CYP proteins.

S.No	Protein ID	Description	Position	Score	MSS	P-value	E-value	Pattern	MSS P (%)
1.	Q0Z7U8_SOR BI Q0Z7U8	Flavonoid 3'- hydroxylase (CYP)	305	135	8.653846e- 01	2.257816e- 10	6.301969e- 05	FGAGRRICA G	86.53
2.	Q4JEG2_SOR BI Q4JEG2	Flavonoid 3'- hydroxylase (CYP)	443	135	8.653846e- 01	2.257816e- 10	6.301969e- 05	FGAGRRICA G	86.53
3.	Q0Z7U7_SOR BI Q0Z7U7	Flavonoid 3'- hydroxylase (CYP)	441	135	8.653846e- 01	2.257816e- 10	6.301969e- 05	FGAGRRICA G	86.53
4.	Q0Z7U9_SOR BI Q0Z7U9	Flavonoid 3'- hydroxylase (CYP)	443	135	8.653846e- 01	2.257816e- 10	6.301969e- 05	FGAGRRICA G	86.53
5.	C71E1_SORBI O48958	Cytochrome P450 71E1 (EC 1.14.13.68) (4- hydroxyphenylacetal dehyde oxime monooxygenase	460	138	8.846154e- 01,	6.709244e- 11	1.872671e- 05	FGAGRRICP G	86.53

Supplementary Table 9. Details of predicted heme binding motif in *Glycin max*. To avoid the false positives, cutoff score used for proteome-wide search through derived weight matrix was ≥ 90 (57.69%).

S.No	Protein ID	Description	Position	Score	MSS	P-value	E-value	Pattern	MSSP (%)
1.	Q2LAJ8_SOYB N Q2LAJ8	Cytochrome P450 monooxygenase CYP76E3.	282	138	8.846154e- 01	3.775245e- 11	2.518100e- 05	FGAGRRICP G	88.46
2.	C71DA_SOYB N O48923	Cytochrome P450 71D10	443	138	8.846154e- 01	3.775245e- 11	2.518100e- 05	FGAGRRICP G	88.46
3.	Q2LAK3_SOYB N Q2LAK3	Cytochrome P450 monooxygenase CYP89H3	310	138	8.846154e- 01	3.775245e- 11	2.518100e- 05	FGAGRRICP G	88.46
4.	A5Y5L3_SOYB N A5Y5L3	Flavonoid 3'5' hydroxylase (CYP)	438	135	8.653846e- 01	1.244284e- 10	8.299411e- 05	FGAGRRIC AG	86.53
5.	Q84JJ4_SOYBN Q84JJ4	Flavonoid 3'- hydroxylase	248	135	8.653846e- 01	1.244284e- 10	8.299411e- 05	FGAGRRIC AG	86.53
6.	Q84U83_SOYB N Q84U83	Flavonoid 3'- hydroxylase	65	135	8.653846e- 01	1.244284e- 10	8.299411e- 05	FGAGRRIC AG	86.53
7.	Q8W3Y5_SOY BN Q8W3Y5	Flavonoid 3'- hydroxylase	438	135	8.653846e- 01	1.244284e- 10	8.299411e- 05	FGAGRRIC AG	86.53
8.	Q84U81_SOYB N Q84U81	Flavonoid 3'- hydroxylase	250	135	8.653846e- 01	1.244284e- 10	8.299411e- 05	FGAGRRIC AG	86.53
9.	A5Y5L4_SOYB N A5Y5L4	Flavonoid 3'5' hydroxylase	438	135	8.653846e- 01	1.244284e- 10	8.299411e- 05	FGAGRRIC AG	86.53
10.	Q6YLS3_SOYB N Q6YLS3	Flavonoid 3', 5'- hydroxylase	437	135	8.653846e- 01	1.244284e- 10	8.299411e- 05	FGAGRRIC AG	86.53
11.	Q2LAL2_SOYB N Q2LAL2	Cytochrome P450 monooxygenase CYP76O2	434	130	8.333333e- 01	3.212254e- 10	2.142583e- 04	FGSGRRICP G	83.33

12.	C82A2_SOYBN O81972	Cytochrome P450 82A2 (P450 CP4).	451	129	8.269231e- 01	5.731104e- 10	3.822663e- 04	FGGRRICP G	82.69
13.	Q2LAK1_SOYB N Q2LAK1	Cytochrome P450 monooxygenase CYP77A3v2	147	128	8.205128e- 01	6.981132e- 10	4.656436e- 04	FGVRRICP G	82.05
14.	C77A3_SOYBN O48928	Cytochrome P450 77A3	443	128	8.205128e- 01	6.981132e- 10	4.656436e- 04	FGVRRICP G	82.05
15.	C71D9_SOYBN O81971	Cytochrome P450 71D9 (P450 CP3).	429	127	8.141026e- 01	1.006429e- 09	6.712913e- 04	FGAGRRMC PG	81.41
16.	C71D8_SOYBN O81974	Cytochrome P450 71D8 (P450 CP7)	436	127	8.141026e- 01	1.006429e- 09	6.712913e- 04	FGAGRRMC PG	81.41
17.	Q2LAK5_SOYB N Q2LAK5	Cytochrome P450 monooxygenase CYP71D54	199	127	8.141026e- 01	1.006429e- 09	6.712913e- 04	FGAGRRMC PG	81.41
18.	Q2LAL4_SOYB N Q2LAL4	Cytochrome P450 monooxygenase CYP83E8	433	126	8.076923e- 01	1.333465e- 09	8.894250e- 04	FGAGRRLC PG	80.76
19.	C98A2_SOYBN O48922	Cytochrome P450 98A2	431	120	7.692308e- 01	5.228881e- 09	3.487679e- 03	FGSGRRVC PG	76.92
20.	C82A4_SOYBN O49859	Cytochrome P450 82A4 (P450 CP9).	454	120	7.692308e- 01	5.228881e- 09	3.487679e- 03	FGSGRRVC PG'	76.92
21.	Q2LAL3_SOYB N Q2LAL3	Cytochrome P450 monooxygenase CYP84A16	449	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRSCP G	76.28
22.	O48924_SOYB N O48924	CYP83D1p.	448	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRMC PA	76.28

23.	Q9M6B9_SOYB N Q9M6B9	Isoflavone synthase 1	439	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRMC PG	76.28
24.	Q9M6D6_SOY BN Q9M6D6	Isoflavone synthase 1	441	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRMC PG	76.28
25.	A6XKY3_SOY BN A6XKY3	Isoflavone synthase 1	441	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRMC PG	76.28
26.	Q9XHP5_SOYB N Q9XHP5	Cytochrome P450 H2O2-dependent urate-degrading peroxidase.	431	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRMC PG	76.28
27.	O48926_SOYB N O48926	CYP93C1p (Isoflavone synthase 2)	441	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRMC PG	76.28
28.	Q9SWR5_SOY BN Q9SWR5	Cytochrome P450 monooxygenase CYP93C1v2p	441	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRMC PG	76.28
29.	Q5DP49_SOYB N Q5DP49	Isoflavone synthase 2	446	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRMC PG	76.28
30.	Q1JV37_SOYB N Q1JV37	Beta-amyrin and sophoradiol 24- hydroxylase	443	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRSCP G	76.28
31.	Q9XHC6_SOY BN Q9XHC6	Cytochrome P450 monooxygenaseC YP93D1	443	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRSCP G	76.28
32.	C93A2_SOYBN Q42799	Cytochrome P450 93A2	432	119	7.628205e- 01	7.520417e- 09	5.016141e- 03	FGSGRRSCP G	76.28
33.	C71A9_SOYBN O81970	Cytochrome P450 71A9 (P450 CP1).	432	118	7.564103e- 01	1.215547e- 08	8.107736e- 03	FGVGRRGC PG	75.64
34.	C78A3_SOYBN O48927	Cytochrome P450 78A3	459	118	7.564103e- 01	1.215547e- 08	8.107736e- 03	FGSGRRAC PG	75.64

35.	O48925_SOYB N O48925	CYP82C1p	462	118	7.564103e-01	1.215547e-08	8.107736e-03	FGSGRRAC PG	75.64
36.	C93A3_SOYBN O81973	Cytochrome P450 93A3 (P450 CP5).	440	118	7.564103e-01	1.215547e-08	8.107736e-03	FGSGRRAC PG	75.64
37.	C93A1_SOYBN Q42798	Cytochrome P450 93A1	439	118	7.564103e-01	1.215547e-08	8.107736e-03	FGSGRRTC PG	75.64
38.	TCMO_SOYBN Q42797	Trans-cinnamate 4- monooxygenase (EC 1.14.13.11)	440	117	7.500000e-01	1.659874e-08	1.107141e-02	FGVGRRSC PG'	75.00
39.	O48918_SOYB N O48918	CYP71A10	444	117	7.500000e-01	1.659874e-08	1.107141e-02	FIGRRC A	75.00
40.	C82A3_SOYBN O49858	Cytochrome P450 82A3 (P450 CP6).	456	117	7.500000e-01	1.659874e-08	1.107141e-02	FGSGRRVC AG	75.00
41.	Q2LAK9_SOYB N Q2LAK9	Cytochrome P450 monooxygenase CYP82E13 (Cytochrome P450 monooxygenase CYP81E10).	432	115	7.371795e-01	2.048221e-08	1.366169e-02	FGMGRRAC PG	73.71
42.	Q2LAL0_SOYB N Q2LAL0	Cytochrome P450 monooxygenase CYP81E11	432	115	7.371795e-01	2.048221e-08	1.366169e-02	FGMGRRAC PG	73.71
43.	Q2LAK0_SOYB N Q2LAK0	Cytochrome P450 monooxygenase CYP701A16	436	108	6.923077e-01	1.203576e-07	8.027890e-02	FGAGKRVC AG	69.23
44.	A5YN45_SOYB N A5YN45	Cytochrome P450 monooxygenase	417	108	6.923077e-01	1.203576e-07	8.027890e-02	FGAGKRVC AG	69.23

45.	Q2LAK6_SOYB N Q2LAK6	CYP701A16 Cytochrome P450 monooxygenase	393	104	6.666667e- 01	2.336797e- 07	1.558651e- 01	FGGGPRICI A	66.66
46.	Q2LAK2_SOYB N Q2LAK2	CYP72A68 Cytochrome P450 monooxygenase	292	101	6.474359e- 01	5.738629e- 07	3.827683e- 01	FGGGPRLC PG	64.74
47.	Q1WCP0_SOY BN Q1WCP0	CYP90A15 Cytochrome P450 monooxygenase	323	94	6.025641e- 01	2.250528e- 06	1.501109e+ 00	FGGGPRKC VG	60.25
48.	C97B2_SOYBN O48921	CYP97C10 Cytochrome P450 97B2	511	94	6.025641e- 01	2.250528e- 06	1.501109e+ 00	FGGGPRKC VG	60.25
49.	Q2LAJ9_SOYB N Q2LAJ9	CYP51G1 Cytochrome P450 monooxygenase	421	93	5.961538e- 01	2.757692e- 06	1.839389e+ 00	FGGGRHGC LG	59.61
50.	Q1WCP1_SOY BN Q1WCP1	CYP94D24 Cytochrome P450 monooxygenase	236	91	5.833333e- 01	4.698946e- 06	3.134211e+ 00	FGSERRLRE S	58.33
51.	Q2LAK7_SOYB N Q2LAK7	CYP72A67 Cytochrome P450 monooxygenase	345	91	5.833333e- 01	4.698946e- 06	3.134211e+ 00	FGWGPRLC IG	58.33
52.	Q2HWE6_SOY BN Q2HWE6	Hypothetical DNA binding protein	175	90	5.769231e- 01	6.050316e- 06	4.035579e+ 00	FGLGQRLP LS	57.69

Supplementary Table 10. Details of predicted heme binding motif in *Brassica napus*. To avoid the false positives, cutoff score used for proteome-wide search through derived weight matrix was ≥ 90 (57.69%). High score proteins belongs to CYP.

S.No	Protein ID	Description	Position	Score	MSS	P-value	E-value	Pattern	MSSP (%)
1.	A1XBC6_BRANA A1XBC6	Flavonoid 3'-hydroxylase (CYP)	435	135	8.653846e-01	1.737243e-10	5.405641e-05	FGAGRRIC AG	86.53
2.	Q9FVC0_BRANA Q9FVC0	Cytochrome p450-dependent monooxygenase	450	119	7.628205e-01	1.001628e-08	3.116686e-03	FGSGRRS CPG	76.28
3.	Q0PW93_BRANA Q0PW93	Ferulate-5-hydroxylase	450	119	7.628205e-01	1.001628e-08	3.116686e-03	FGSGRRS CPG	76.28
4.	Q9FVB9_BRANA Q9FVB9	Cytochrome p450-dependent monooxygenase	450	119	7.628205e-01	1.001628e-08	3.116686e-03	FGSGRRS CPG	76.28
5.	A5GZU7_BRANA A5GZU7	Cinnamate 4-hydroxylase isoform 2 (CYP)	439	117	7.500000e-01	2.201003e-08	6.848684e-03	FGVGRRS CPG	75.00
6.	A5GZU5_BRANA A5GZU5	Cinnamate 4-hydroxylase isoform 1	439	117	7.500000e-01	2.201003e-08	6.848684e-03	FGVGRRS CPG	75.00
7.	Q9FVB8_BRANA Q9FVB8	Cytochrome p450-dependent monooxygenase	443	101	6.474359e-01	7.261404e-07	2.259473e-01	FGSGRKS CPG	64.74
8.	Q9STD9_BRANA Q9STD9	Cellulase (Endo-1,4-beta-d-glucanase) (EC 3.2.1.4).	303	90	5.769231e-01	7.477900e-06	2.326838e+00	FGRTRRG RYS	57.69

Supplementary Table 11. Details of predicted heme binding motif in *Brassica oleracea*. To avoid the false positives, cutoff score used for proteome-wide search through derived weight matrix was ≥ 78 (50%). No high score motif found in CYPs.

S.No	Protein ID	Description	Position	Score	MSS	P-value	E-value	Pattern	MSSP (%)
1.	Q25BL3_BRAOL Q25BL3	Putative uncharacterized protein	254	88	5.641026e-01	1.341351e-05	3.324766e+00	FGLGRRT RSV	56.41
2.	Q84KT7_BRAOL Q84KT7	S-locus protein 11	14	88	5.641026e-01	1.341351e-05	3.324766e+00	KGTFRRR CGS	56.41
3.	Q2A9I4_BRAOL Q2A9I4	S-locus linked 3 (SLL3) protein, putative	254	88	5.641026e-01	1.341351e-05	3.324766e+00	FGLGRRT RSV	56.41
4.	Q2A9F7_BRAOL Q2A9F7	SLL3 ORF2 protein, putative	254	88	5.641026e-01	1.341351e-05	3.324766e+00	FGLGRRT RSV	56.41
5.	Q9ATA2_BRAOL Q9ATA2	Disease resistance-like protein	75	85	5.448718e-01	1.781478e-05	4.415695e+00	FGPGSRIV VT	54.48
6.	Q9ATA3_BRAOL Q9ATA3	Disease resistance-like protein	48	85	5.448718e-01	1.781478e-05	4.415695e+00	FGPGSRIV VT	54.48
7.	Q25BK9_BRAOL Q25BK9	Putative uncharacterized protein	141	84	5.384615e-01	2.262181e-05	5.607200e+00	FGANNRV SRS	53.84
8.	Q84UZ9_BRAOL Q84UZ9	Dihydroflavonol 4-reductase	245	83	5.320513e-01	3.148964e-05	7.805243e+00	AAKGRYI CSS	53.20
9.	Q2A9F0_BRAOL Q2A9F0	Putative uncharacterized protein	315	83	5.320513e-01	3.148964e-05	7.805243e+00	GKAGRRS EES	53.20
10.	Q9FQ62_BRAOL Q9FQ62	Acid invertase (EC 3.2.1.26)	394	83	5.320513e-01	3.148964e-05	7.805243e+00	QNKGRRI LWS	53.20

11.	Q25BK1_BRAOL Q25BK1	Putative uncharacterized protein	104	83	5.320513e- 01	3.148964e- 05	7.80524 3e+00	FGSGVLA GPS	53.20
12.	Q9LEP5_BRAOL Q9LEP5	Putative uncharacterized protein RGL-ld	90	83	5.320513e- 01	3.148964e- 05	7.80524 3e+00	FGPGSRIII T	53.20
13.	Q2A9U4_BRAOL Q2A9U4	Putative uncharacterized protein	339	82	5.256410e- 01	4.178793e- 05	1.03578 5e+01	GGAGARA RNS	52.56
14.	Q2A9X5_BRAOL Q2A9X5	Putative uncharacterized protein	176	81	5.192308e- 01	4.791078e- 05	1.18755 0e+01	FGAGRGY GRG	51.92
15.	Q2A9S4_BRAOL Q2A9S4	Putative uncharacterized protein	64	80	5.128205e- 01	5.321827e- 05	1.31910 5e+01	FGFGDAT APS	51.12
16.	Q2A9X4_BRAOL Q2A9X4	Myb transcription factor, putative	10	80	5.128205e- 01	5.321827e- 05	1.31910 5e+01	NGHNSRT CPS	51.12
17.	Q2A970_BRAOL Q2A970	Trehalose-6- phosphate synthase, putative	560	79	5.064103e- 01	5.809988e- 05	1.44010 4e+01	FGLGFRV VAL	50.64
18.	Q5FAL5_BRAOL Q5FAL5	Putative 2- isopropylmalate synthase	581	78	5.000000e- 01	6.310464e- 05	1.56415 6e+01	TGAGMDI VVS	50

Supplementary Table 12. Details of predicted heme binding motif in *Solanum lycopersicum*. To avoid the false positives, cutoff score used for proteome-wide search through derived weight matrix was ≥ 91 (58.33%). High score motif found in CYPs.

S.No	Protein ID	Description	Position	Score	MSS	P-value	E-value	Pattern	MSSP (%)
1.	A0ZS62_SOLLC A0ZS62	Cytochrome P450	420	100	6.410256e-01	8.965322e-07	6.226712e-01	FGGGSRC CPG	64.10
2.	C85A3_SOLLC Q50LE0	Cytochrome P450 85A3 (C6-oxidase)	409	100	6.410256e-01	8.965322e-07	6.226712e-01	FGGGTRL CPG	64.10
3.	Q0NZZ1_SOLLC Q0NZZ1	6-deoxocastasterone oxidase	406	100	6.410256e-01	8.965322e-07	6.226712e-01	FGGGTRQ CPG	64.10
4.	Q0NZZ0_SOLLC Q0NZZ0	6-deoxocastasterone oxidase	406	100	6.410256e-01	8.965322e-07	6.226712e-01	FGGGTRQ CPG	64.10
5.	C85A1_SOLLC Q43147	Cytochrome P450 85A1 (C6-oxidase) (Dwarf protein)	406	100	6.410256e-01	8.965322e-07	6.226712e-01	FGGGTRQ CPG	64.10
6.	Q0NZY6_SOLLC Q0NZY6	6-deoxocastasterone oxidase	406	100	6.410256e-01	8.965322e-07	6.226712e-01	FGGGTRQ CPG	64.10
7.	A0ZS63_SOLLC A0ZS63	Cytochrome P450	408	98	6.282051e-01	1.200332e-06	8.336701e-01	FGGGPRL CAG	62.82
8.	CB12_SOLLC P10708	Chlorophyll a-b binding protein 7, chloroplast precursor (LHCI type II CAB-7)	33	93	5.961538e-01	2.861674e-06	1.987527e+00	FLGGRRL RVS	59.61
9.	Q9M4X2_SOLLC	Putative cytochrome	438	92	5.897436e-	3.626789e-	2.51892	FGWGPR	58.97

	Q9M4X2	P450			01	06	4e+00	MCIG	
10.	Q1ELT8_SOLLC Q1ELT8	Red chlorophyll catabolite reductase	25	92	5.897436e- 01	3.626789e- 06	2.51892 4e+00	FSVGKRF CCS	58.97
11.	Q0X087_SOLLC Q0X087	Castasterone 26- hydroxylase	455	91	5.833333e- 01	4.854108e- 06	3.37133 8e+00	FGLGARR CVG	58.33
12.	O24019_SOLLC O24019	GTP cyclohydrolase II / 3,4-dihydroxy- 2-butanone-4- phosphate synthase	389	91	5.833333e- 01	4.854108e- 06	3.37133 8e+00	FGSARCD CGS	58.33
13.	GRW10_SOLLC Q01157	Glycine-rich cell wall structural protein (Clone W10-1)	33	91	5.833333e- 01	4.854108e- 06	3.37133 8e+00	RGGGGRK CCS	58.33
14.	Q02021_SOLLC Q02021	Glycine-rich protein	122	91	5.833333e- 01	4.854108e- 06	3.37133 8e+00	RGGGGRK CCS	58.33
15.	Q02022_SOLLC Q02022	Glycine-rich protein	110	91	5.833333e- 01	4.854108e- 06	3.37133 8e+00	RGGGGRK CCS	58.33
16.	Q02023_SOLLC Q02023	Glycine-rich protein	26	91	5.833333e- 01	4.854108e- 06	3.37133 8e+00	RGGGGRK CCS	58.33
17.	Q02024_SOLLC Q02024	Glycine-rich protein	89	91	5.833333e- 01	4.854108e- 06	3.37133 8e+00	RGGGGRK CCS	58.33
