Supplementary Information

for

Identification and characterization of a rhamnosyltransferase involved in rutin biosynthesis in *Fagopyrum esculentum* (common buckwheat)

Eiki Koja[1], Soichiro Ohata[1], Yoshinori Maruyama[1], Hideyuki Suzuki[2], Makoto Shimosaka[1], Goro Taguchi[1,*]

[1]*Department of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, 3-15-1 Tokida, Ueda 386-8567, Japan.* [2]*Department of Research and Development, Kazusa DNA Research Institute, 2-6-7 Kazusa Kamatari, Kisarazu 292-0818, Japan.*

[*]Corresponding author. e-mail : gtagtag@shinshu-u.ac.jp

Table S1. PCR primers used in this study.

Primer	Sequence
SFF-Fw1	5'-TCTTTCTTYTCNGCNCCNGGNAACAT-3'
GDQ-Rv1	5'-TTRGARTTNAARAAYTGRTCNCC-3'
M13-FL	5'-GTTTTCCCAGTCACGACGTTGTA-3'
M13-Rv	5'-CAGGAAACAGCTATGACCAT-3'
FeRhaT-F1	5'-cacccatatgATGGGAACCCAAGCAAAC-3'
FeRhaT-R1	5'-ctcgagTTAATTACCAACCAAAGCCTTAAC-3'
FeRhaT-F2	5'-caccATGGGAACCCAAGCAAAC-3'
FeRhaT-324F	5'-CACCCAACTCAAACCCGA-3'
FeRhaT-548F	5'-AAGACATTCCAGGCGAAGGAC-3'
FeRhaT-653R	5'-CGATTGCATCACATCCTTGG-3'
FeRhaT-838R	5'-CGGATCCGAAGGAGCATAG-3'
FeGAPDH-543F	5'-CGTTGAGGGCCTAATGACCA-3'
FeGAPDH-689R	5'-TTGCCAACAGCCTTAGCAGC-3'
FeCHI-215F	5'-TCACGGCGATCGGAATTTAC-3'
FeCHI-343R	5'-CGAATTGACCGGTGACAACA-3'
FeCHS-221F	5'-CAAGCGCATGTGTGACAAGTC-3'
FeCHS-323R	5'-GGATGGAGCCATGTAAGCACA-3'
AtRhm2-Fw	5'-ggaattcATGGATGATACTACGTATAAGCC-3'
AtRhm2-Rv	5'-ccgctcgagGGTTCTCTTGTTTGGTTCA-3'
35S-1	5'-GCTCCTACAAATGCCATCA-3'
NtActin-Fw	5'-GATTGGAATGGAATGGAAGCTG-3'
NtActin-Rv	5'-CCTCCAATCCAAACACT-3'

Table S2. List of uridine diphosphate (UDP)-sugar dependent glycosyltransferases (UGTs) used in the phylogenetic analysis.

Name	Function	Organism	Accession No.[a]
FeF3G6″RhaT (UGT79A8)	Flavonol 3-*O*-glucoside 6″-*O*-glucosyltransferase	*Fagopyrum esculentum*	LC312144
LeABRT2	Delphinidin 3-*O*-glucoside 6″-*O*-rhamnosyltransferase	*Lobelia erinus*	LC131336
LeABRT4	Delphinidin 3-*O*-glucoside 6″-*O*-rhamnosyltransferase	*Lobelia erinus*	LC131337
PhRT (UGT79A1)	Anthocyanidin 3-*O*-glucoside 6″-*O*-rhamnosyltransferase	*Petunia x hybrida*	X71059
Cm1,2RhaT	Flavanone 7-*O*-glucoside 2″-*O*-rhamnosyltransferase	*Citrus maxima*	AY048882
Cm1,6RhaT	Flavanone 7-*O*-glucoside 6″-*O*-rhamnosyltransferase	*Citrus maxima*	LC057678
Cs1,6RhaT	Flavanone 7-*O*-glucoside 6″-*O*-rhamnosyltransferase	*Citrus sinensis*	DQ119035
GmF3G2″GT (UGT79B30)	Flavonol 3-*O*-glucoside 2″-*O*-glucosyltransferase	*Glycine max*	LC017844
GmSGT3 (UGT91H4)	Soyasaponin III 2″-*O*-rhamnosyltransferase	*Glycine max*	AB473731
CaUGT3 (UGT94E3)	Flavonoid glucoside 6″-*O*-glucosyltransferase	*Catharanthus roseus*	AB443870
GjUGT9 (UGT94E5)	Crocetin glucoside 6′-*O*-glucosyltransferase	*Gardenia jasminoides*	AB555739
SiSG6′GT (UGT89D1)	Sesaminol 2-*O*-glucoside 6′-*O*-glucosyltransferase	*Sesamum indicum*	AB333799
BpUGAT (UGT94B1)	Anthocyanin 3-*O*-glucoside 2″-*O*-glucuronosyltransferase	*Bellis perennis*	AB190262
IpA3G2″GT (UGT79B16)	Anthocyanidin 3-*O*-glucoside 2″-*O*-glucosyltransferase	*Ipomoea purpurea*	AB192315
GmF3G6″RT (UGT79A6)	Flavonol 3-*O*-glucoside 6″-*O*-rhamnosyltransferase	*Glycine max*	AB828193
GmF3G6″GT (UGT79A7)	Flavonol 3-*O*-glucoside 6″-*O*-glucosyltransferase	*Glycine max*	LC126028
AtA3RhaT (UGT78D1)	Anthocyanin 3-*O*-rhamnosyltransferase	*Arabidopsis thaliana*	NM_102790
AtF3G2″GT (UGT79B6)	Flavonoid 3-*O*-glucoside 2″-*O*-glucosyltransferase	*Arabidopsis thaliana*	NM_124780
AtF3G2″XylT (UGT79B1)	Anthocyanin 3-*O*-glucoside 2″-*O*-xylosyl-transferase	*Arabidopsis thaliana*	NM_124785
AtF7RhaT (UGT89C1)	Flavonol 7-*O*-rhamnosyltransferase	*Arabidopsis thaliana*	NM_100480

[a] Accession No. for gene sequences.

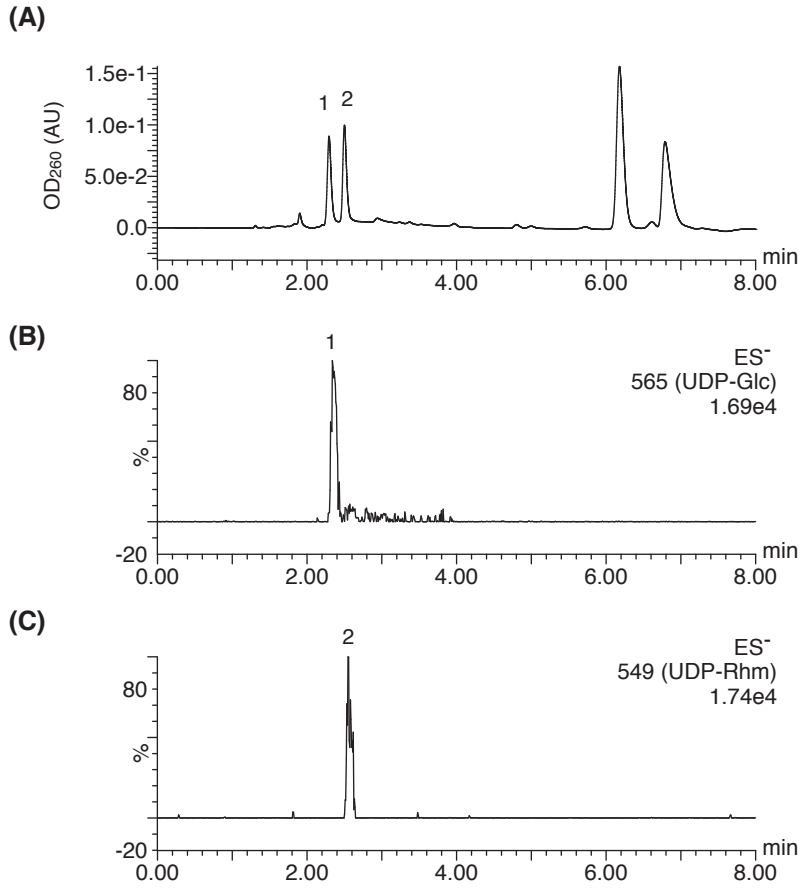


Figure S1. HPLC analysis of the UDP-rhamnose solution synthesized by the enzyme reaction. UDP-rhamnose was synthesized from UDP-glucose via an enzymatic reaction using Arabidopsis Rhm2 and was analyzed by HPLC as described in the Materials and Methods Section. The eluates were monitored at 260 nm using a diode array detector (A), the negative electron-spray ionization (ES^-) MS at m/z = 565 for UDP-glucose (B), and at m/z = 549 for UDP-rhamnose (C). The retention time of the MS peaks was delayed by about 0.08 min over that of the diode array. Peak identification: 1, UDP-glucose; 2, UDP-rhamnose.

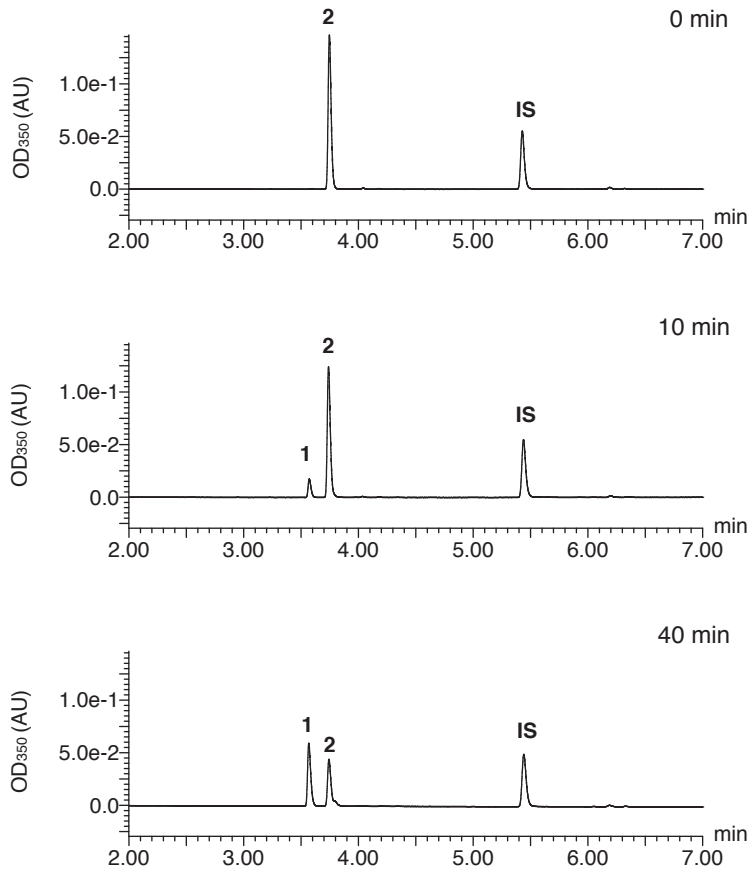


Figure S2. Rhamnosyltransferase activity in the cell-free extract of buckwheat etiolated cotyledon. Cell-free extract of buckwheat etiolated cotyledon was incubated with isoquercitrin (quercetin 3-*O*-glucoside) and UDP-rhamnose solution at 30 °C for 0, 10, and 40 min, and analyzed by HPLC with a 100 mm ODS column as described in the Materials and Methods Section. The eluates were monitored at 350 nm using a diode array detector. Peak identification: 1, rutin; 2, isoquercitrin; IS, quercetin added to the samples as an internal standard.

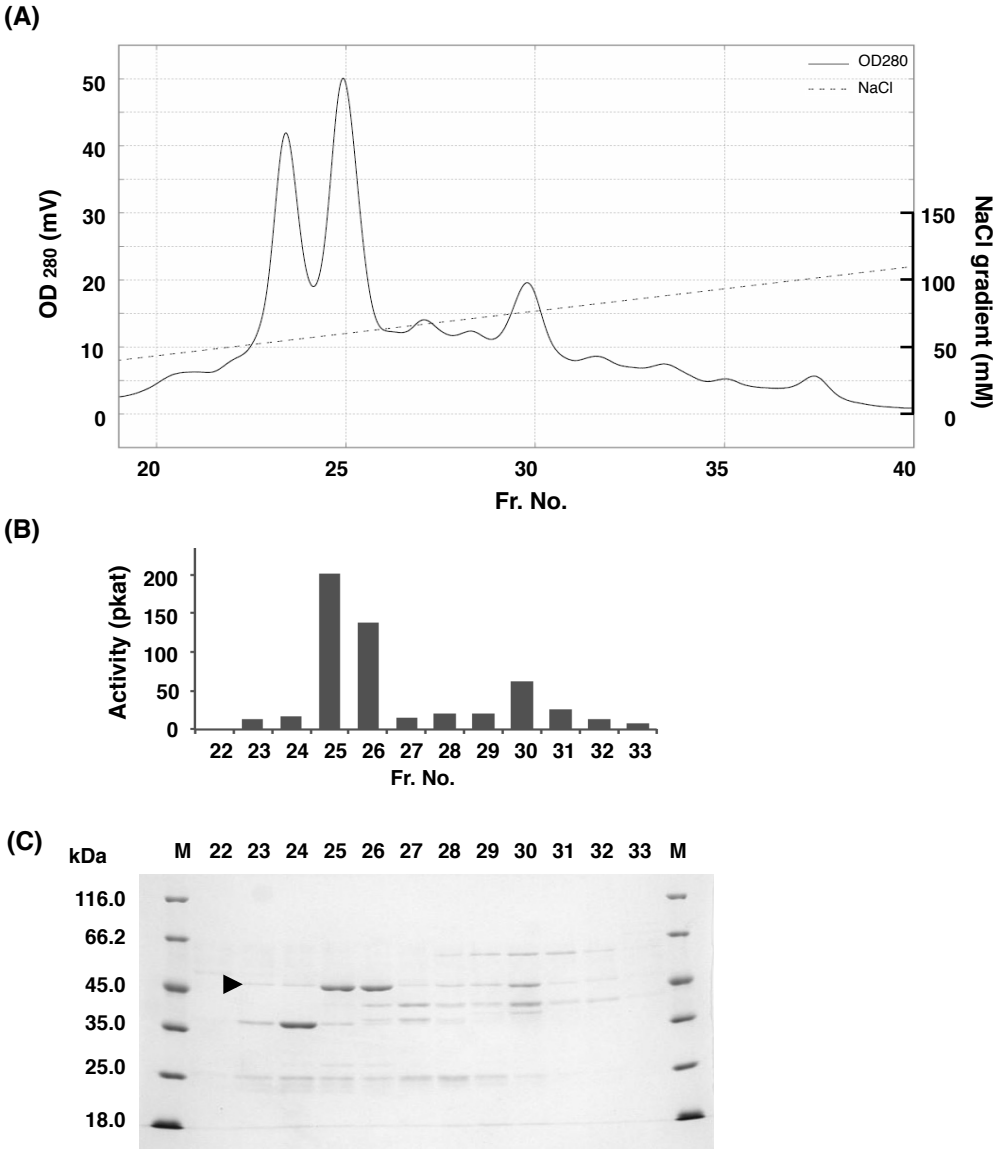


Figure S3. Purification of buckwheat rhamnosyltransferase on Mono Q anion exchange chromatography. (A) Chromatogram of rhamnosyltransferase from buckwheat seedlings on Mono Q. The solid and dashed lines represent the protein concentration (OD_{280}) and NaCl concentration, respectively. (B, C) Rhamnosyltransferase activity (B) and SDS-PAGE analysis (C) of the purified fractions after Mono Q separation, respectively. The numbers indicate the fraction number on Mono Q shown in (A). Arrowhead in (C) indicates the purified rhamnosyltransferase. Lane M: standard proteins.

```
   1  ATGGGAACCCAAGCAAACACAACCGATCTTCACATAGCCGTATTTCCCTACTTCGCCTTCGGCCACATCAACCCATTCGTTCACATCTCA
        M  G  T  Q  A  N  T  T  D  L  H  I  A  V  F  P  Y  F  A  F  G  H  I  N  P  F  V  H  I  S

  91  AACAAGCTCGCCTCCCATGGAATCAAGATCTCCTTCTTCTCCGCTCCAGGGAACATCCCAAGAATCAAATCGTCACTTTCCACCTCACCC
        N  K  L  A  S  H  G  I  K  I  S  F  F  S  A  P  G  N  I  P  R  I  K  S  S  L  S  T  S  P

 181  TTGATCTCAATCGTACCGCTCACCTTCCCCCACGTCGACGGCCTCCCCGCCGGCTTCGAAAGCACTGCTGACATCACTCCCGCCATTGCT
        L  I  S  I  V  P  L  T  F  P  H  V  D  G  L  P  A  G  F  E  S  T  A  D  I  T  P  A  I  A

 271  GAGCTTCTCAAGGTCGCTCTTGACAAAATGCAGCCTCAAATTCGTTCTCTGCTCACCCAACTCAAACCCGACGTCGTTTTCTTCGACTTC
        E  L  L  K  V  A  L  D  K  M  Q  P  Q  I  R  S  L  L  T  Q  L  K  P  D  V  V  F  F  D  F

 361  GCTCAGAATTGGATCCCTTCTCTTGCCTCCGAACTTGGGATTAAGACTGTTATGTTTTCCGTCTTCTCTCTTATCTCCAACTCTTATTTA
        A  Q  N  W  I  P  S  L  A  S  E  L  G  I  K  T  V  M  F  S  V  F  S  L  I  S  N  S  Y  L

 451  ATGACGCCGGCGAGACTTTCCTCCGACGAGATTCCGACCATTGAAGAGCTCAAGAAACCTCCTCAAGGCTATCCCAACCCCGACCTCTCC
        M  T  P  A  R  L  S  S  D  E  I  P  T  I  E  E  L  K  K  P  P  Q  G  Y  P  N  P  D  L  S

 541  CTGAAGACATTCCAGGCGAAGGACTTGTTGTATCCGTTCAGACGGTTCAACGGCGGTCCATCGGCGCTGGAGCGGAACTACGCTGGAATC
        L  K  T  F  Q  A  K  D  L  L  Y  P  F  R  R  F  N  G  G  P  S  A  L  E  R  N  Y  A  G  I

 631  CAAGGATGTGATGCAATCGCTTACAAGTCTTGTCACGAGATGGAAGGTCCATACTGGAGCTACTTCAAGAAAGTCATCGGAAAGCCAATC
        Q  G  C  D  A  I  A  Y  K  S  C  H  E  M  E  G  P  Y  W  S  Y  F  K  K  V  I  G  K  P  I

 721  ATAATGGCCGGAATTCCGATCCCGGAAACGTCTTCCTCCGGCGACCTCGACAGCAACTGGGCAACATGGCTAGCAAAATTCCCACCAAAA
        I  M  A  G  I  P  I  P  E  T  S  S  S  G  D  L  D  S  N  W  A  T  W  L  A  K  F  P  P  K

 811  TCAGTTACTCTATGCTCCTTCGGATCCGAAACGTTTCTCACCGACGTCCAAGTCCAAGAGCTTGCTCTTGGACTTGAACTCACAGAGCTT
        S  V  T  L  C  S  F  G  S  E  T  F  L  T  D  V  Q  V  Q  E  L  A  L  G  L  E  L  T  E  L

 901  CCATTTCTAATGGTACTGAGCTCCAATGGCTTCGATCAAGAAAGACTGAACAAAATCCTCCCTGAGGGGTTCCTGGAGCGGGTTAAAGAT
        P  F  L  M  V  L  S  S  N  G  F  D  Q  E  R  L  N  K  I  L  P  E  G  F  L  E  R  V  K  D

 991  AGAGGCTTGATTCATATCGGTTGGGTGCCACAGCAGAAGATTATGGCTCATGAGAATGTGGGTTGTTATGTTAATCATGCTGGGTTTGGA
        R  G  L  I  H  I  G  W  V  P  Q  Q  K  I  M  A  H  E  N  V  G  C  Y  V  N  H  A  G  F  G

1081  TCTGTGATTGAAGCCATTGTTACTGATTGTCAGCTGGTTTTGCTTCCATTTAAAGGCGACCAGTTCTTGAACTCGAAGCTGTTGAGTCTG
        S  V  I  E  A  I  V  T  D  C  Q  L  V  L  L  P  F  K  G  D  Q  F  L  N  S  K  L  L  S  L

1171  GACATGAAGGTTGGGGTGGAGGTAAATAGGAGAGATGAAGATGGGCATTTTGGGAAAGAGGATATATTTGAGGCAGTGAGGATTGTTACA
        D  M  K  V  G  V  E  V  N  R  R  D  E  D  G  H  F  G  K  E  D  I  F  E  A  V  R  I  V  T

1261  GTGGATGGTGATAAAGAGCCTGGGAAGAAGATTAGAGGTAATCTTGTGAAGTGGAAGGAGTTGCTGATGAACAAAGAGTTTGAAGAGAAG
        V  D  G  D  K  E  P  G  K  K  I  R  G  N  L  V  K  W  K  E  L  L  M  N  K  E  F  E  E  K

1351  TATGTTCTTGAATTGGTTAAGGAAGTTAAGGCTTTGGTTGGTAATTAA   1398
        Y  V  L  E  L  V  K  E  V  K  A  L  V  G  N  *
```

Figure S4. Nucleotides and deduced amino acid sequence of *FeF3G6"RhaT*.

The peptide sequences obtained by LC-MS/MS analysis of rhamnosyltransferase purified from buckwheat seedlings are highlighted in green.

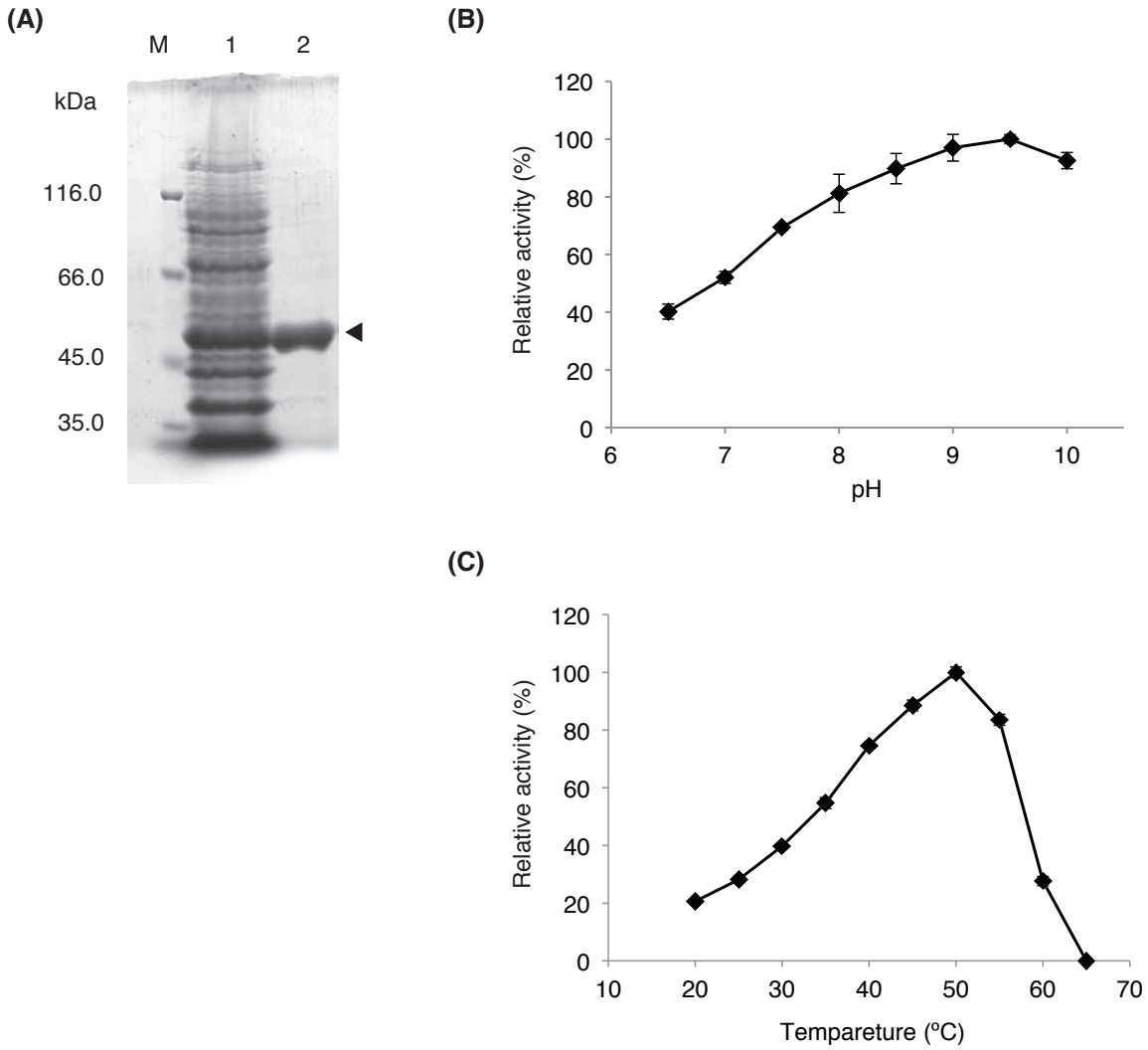
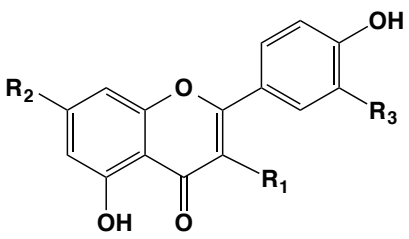


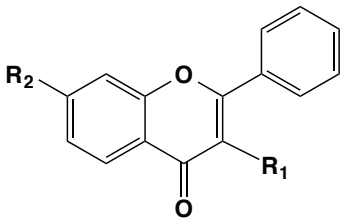
Figure S5. Properties of recombinant FeF3G6″RhaT.

(A) SDS-PAGE analysis of the recombinant FeF3G6″RhaT. Proteins were separated on a 10% SDS-polyacrylamide gel. Lanes 1 and 2, Crude extract and an affinity-purified fraction of the recombinant FeF3G6″RhaT enzyme, respectively; lane M, standard proteins. (B, C) The pH preference (B) and optimal temperature (C) for the FeF3G6″RhaT reaction. The reactions were examined using UDP-rhamnose and quercetin 3-*O*-glucoside as substrates, as described in the Materials and Methods Section. Relative activity is given as the average ± SD (n = 3) with maximum activity levels assumed to be 100%.

Flavones and flavonols

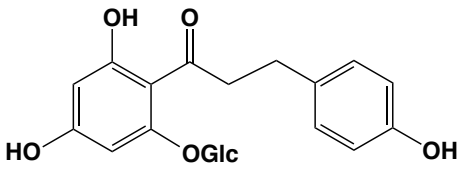


Quercetin	R_1=R_2=R_3=OH
Kaempferol	R_1=R_2=OH, R_3=H
Quercetin 3-*O*-glucoside	R_1=OGlc, R_2=R_3=OH
Quercetin 3-*O*-galactoside	R_1=OGal, R_2=R_3=OH
Kaempferol 3-*O*-glucoside	R_1=OGlc, R_2=OH, R_3=H
Quercetin 7-*O*-glucoside	R_1=R_3=OH, R_2=OGlc
Kaempferol 7-*O*-glucoside	R_1=OH, R_2=OGlc, R_3=H
Kaempferol 3,7-di-*O*-glucoside	R_1=R_2=OGlc, R_3=H

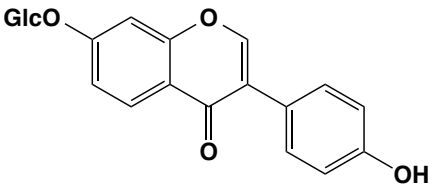


3-Hydroxyflavone glucoside	R_1=OGlc, R_2=H
7-Hydroxyflavone glucoside	R_1=H, R_2=OGlc

Dihydrochalcone and isoflavone

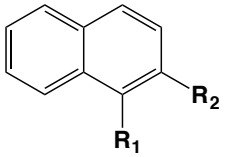


Phloridzin

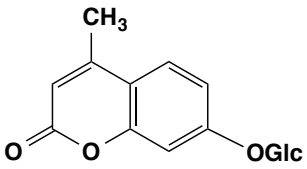


Daidzin

Others



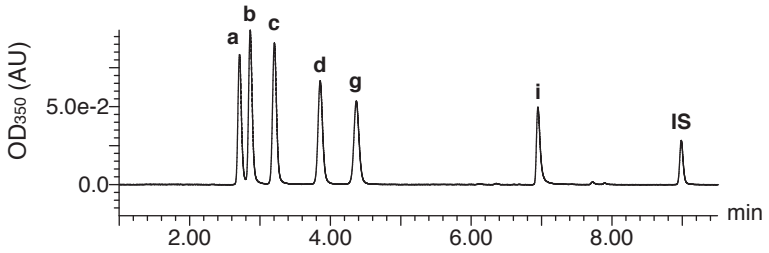
1-Naphthol glucoside	R_1=OGlc, R_2=H
2-Naphthol glucoside	R_1=H, R_2=OGlc



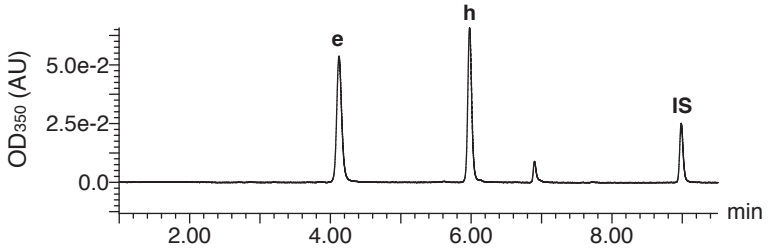
4-Methylumbelliferone glucoside

Figure S6. Structures of the substrates used in this study.

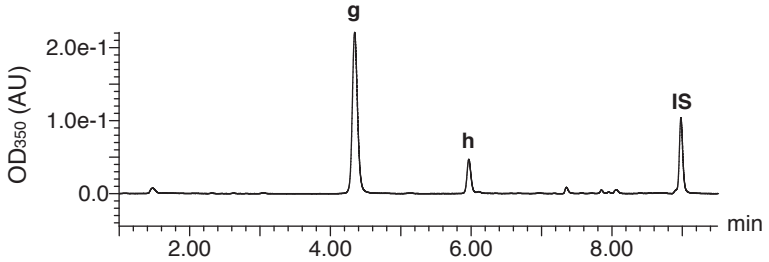
(A) Flavonoid standards 1



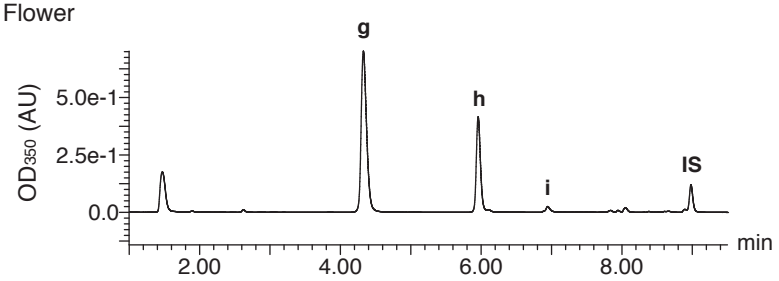
(B) Flavonoid standards 2



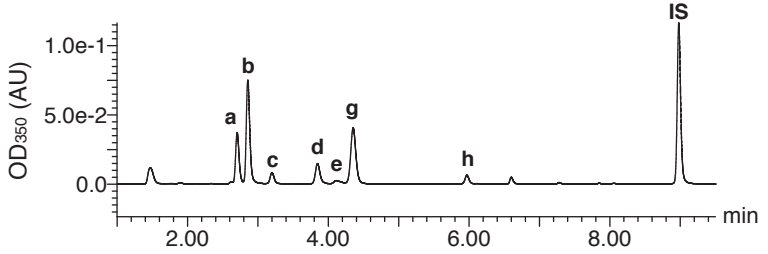
(C) Floral bud



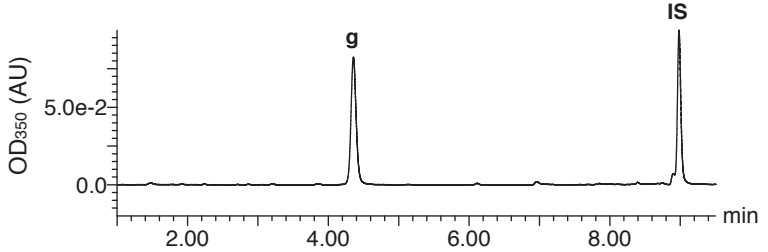
(D) Flower



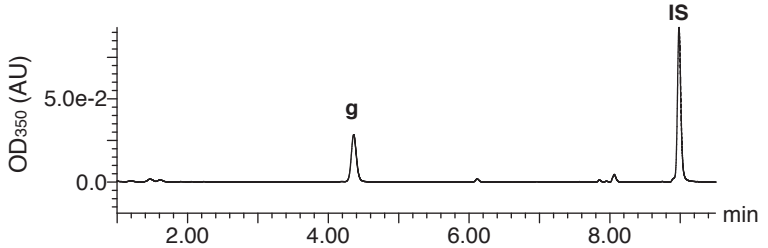
(E) Immature seed



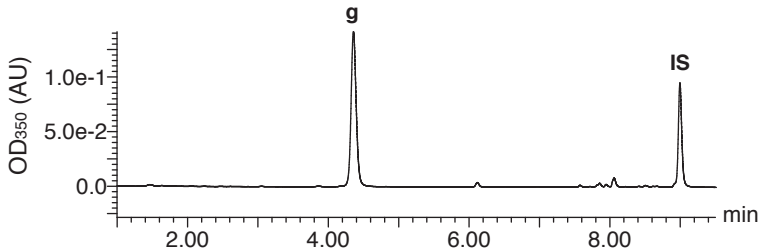
(F) Leaf bud



(G) Young leaf



(H) Matured leaf



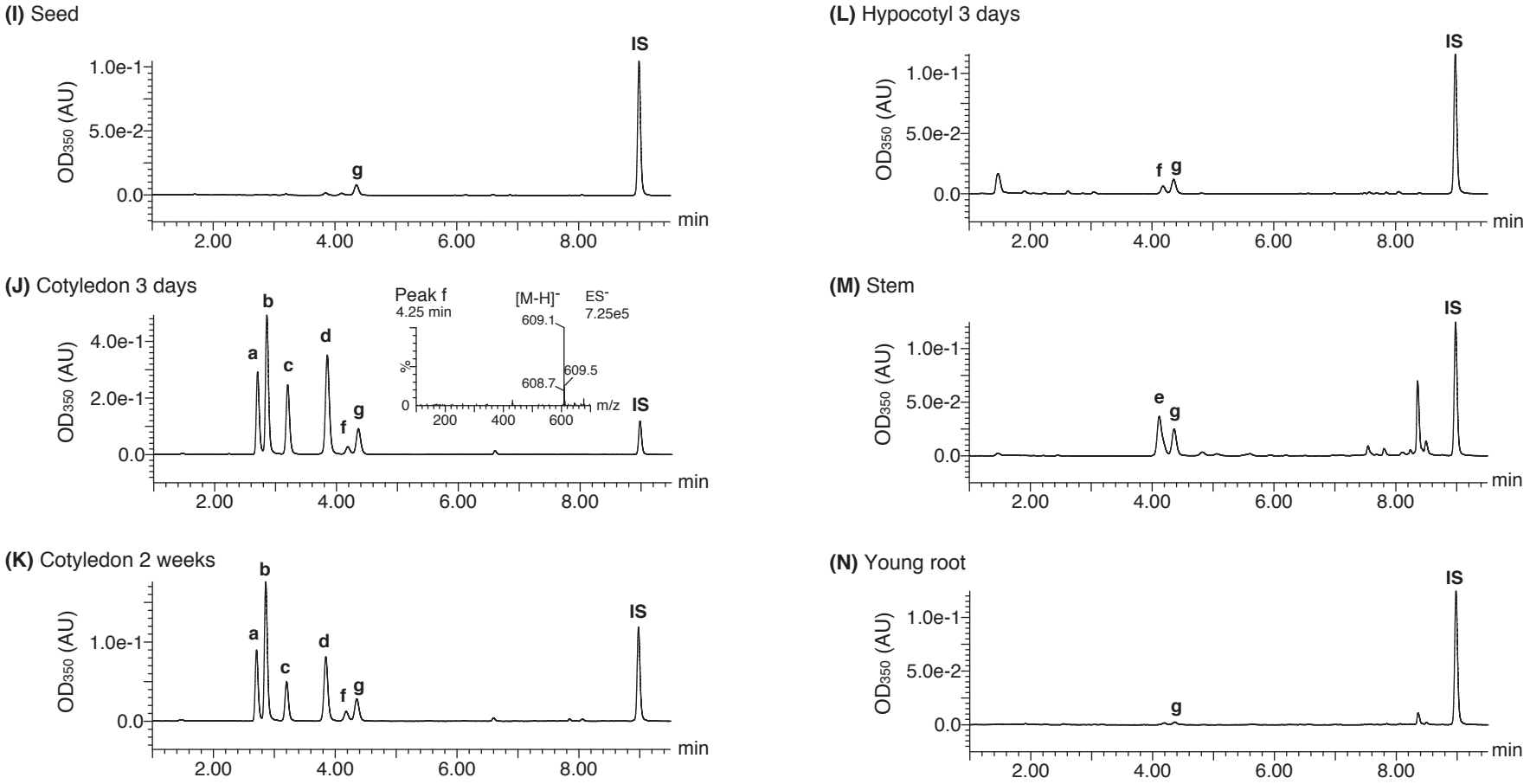


Figure S7. Distribution of flavonoids in the organs of common buckwheat.

Chromatograms reflect the following conditions: (A, B) authentic compounds of flavonoids; methanol extract of the floral bud (C), flower including bud (D), immature seed (E), leaf bud of the primary leaf (F), primary leaf after development (G), matured leaf (H), seed (I), cotyledons [3 days (J) and 2 weeks (K) after sowing], hypocotyl 3 days after sowing (L), stem (red part near the root) (M), and young root from seedling (N). The eluates were detected by measuring the absorbance at 350 nm. The negative electron-splay ionization (ES^-) MS spectra corresponding to the peak f is shown. The retention time of MS peaks was delayed by about 0.08 min compared with that of the diode array. Peak identifications: a, orientin; b, isoorientin; c, vitexin; d, isovitexin; e, quercetin 3-*O*-galactoside; f, a compound predicted to be quercetin 3-*O*-robinobioside; g, rutin; h, quercitrin; i, quercetin; IS, chrysin added to the samples as an internal standard.

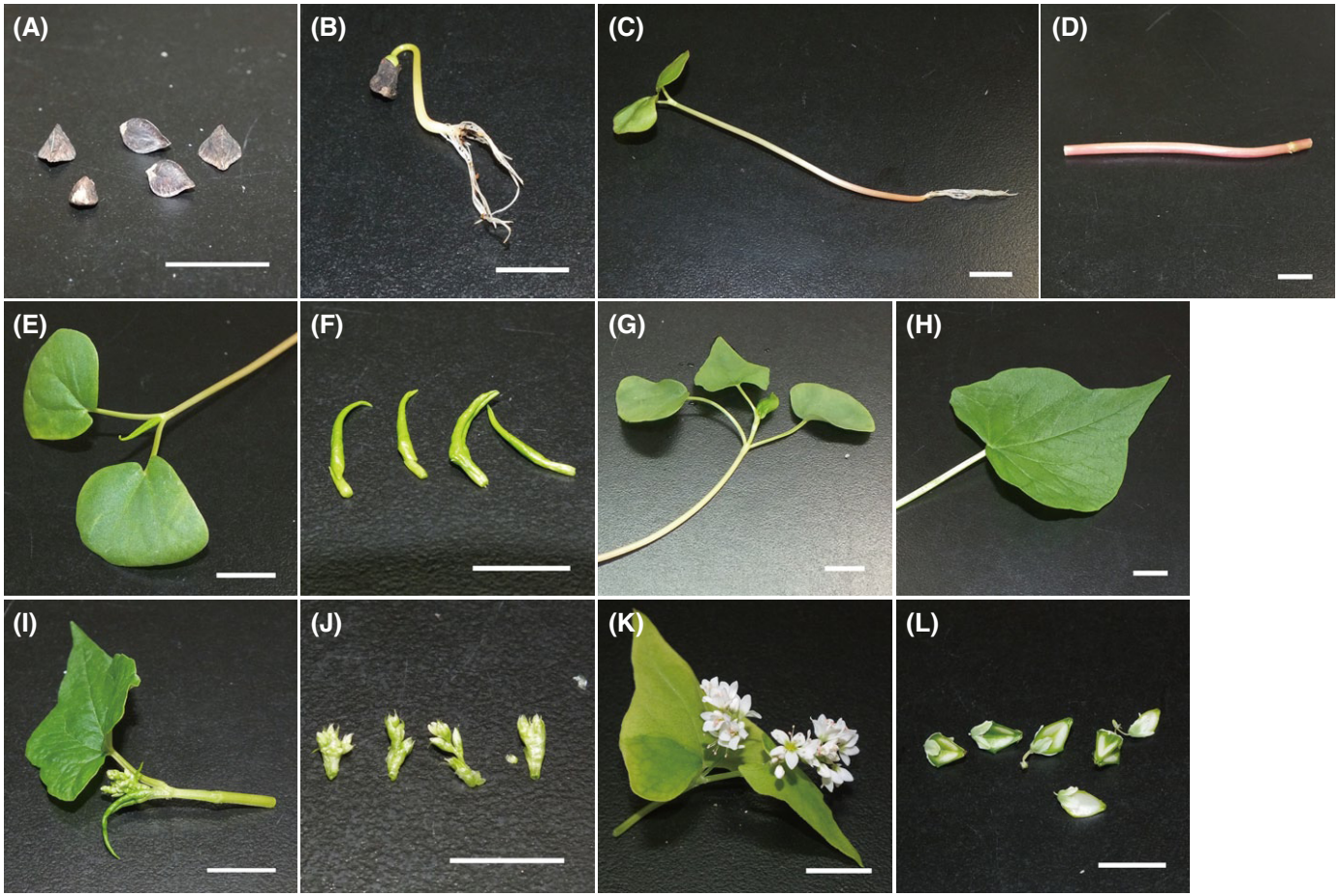


Figure S8. Images of common buckwheat organs used in this study.

The panels show the common buckwheat organs in several developmental stages: seeds (A); 3 day-old seedling for cotyledon and hypocotyl preparation (B); 1-week-old seedling for young root preparation (C); stem (D); 2-week-old seedling for cotyledon and leaf bud (E), and separated leaf buds (F); 3-week-old seedling for primary leaf (G); matured leaf (H); floral bud with leaves (I) and separated floral buds (J); flowering stage including flowers and buds (K); immature seeds (L). White scale bar indicates 1 cm.

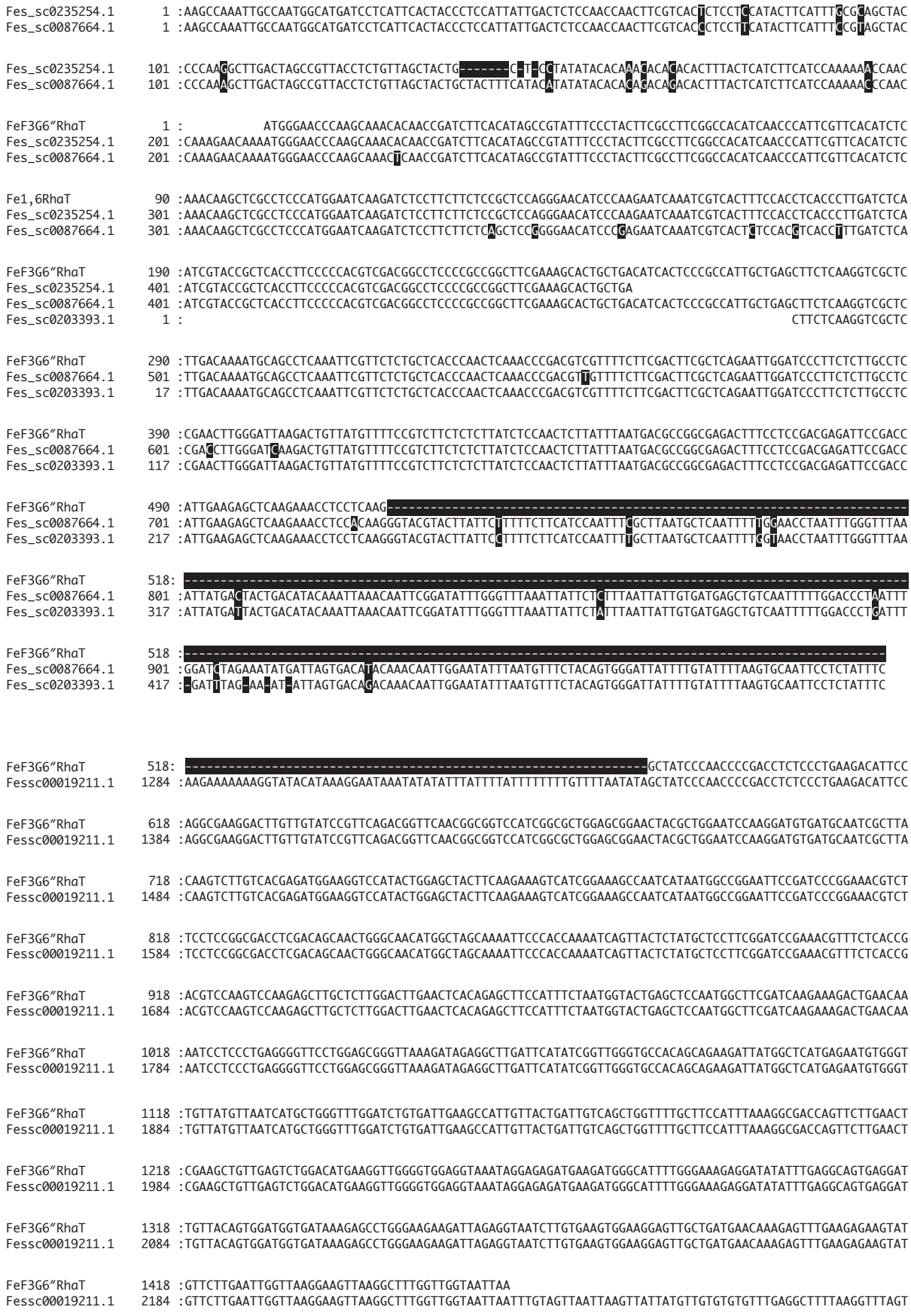


Figure S9. Comparison of the DNA sequences of *FeF3G6"RhaT* and related sequences found in the draft genome database of common buckwheat.

Black shading shows the different nucleotides among the sequences. Fessc00019211.1 is shown only by the relevant part, and any homologous sequence was identified in the omitted 1.3 kb part corresponding to the intron.