Available online at www.sciencedirect.com
ScienceDirect
Phytochemistry 69 (2008) 99–111
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Metabolic profiling of flavonoids in *Lotus japonicus* using liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry

Hideyuki Suzuki [a], Ryosuke Sasaki [a], Yoshiyuki Ogata [a], Yukiko Nakamura [b,d,1], Nozomu Sakurai [a], Mariko Kitajima [c], Hiromitsu Takayama [c], Shigehiko Kanaya [b], Koh Aoki [a,*], Daisuke Shibata [a], Kazuki Saito [c,*]

[a] *Kazusa DNA Research Institute, Kazusa-Kamatari 2-6-7, Kisarazu 292-0818, Japan*
[b] *Graduate School of Information Science, Nara Institute of Science and Technology, Takayama 8916-5, Ikoma, Nara 630-0101, Japan*
[c] *Graduate School of Pharmaceutical Sciences, Chiba University, Yayoi-cho 1-33, Inage-ku, Chiba 263-8522, Japan*
[d] *Ehime Women's College, Baba, Uwajima 798-0025, Japan*

Received 8 March 2007; received in revised form 21 May 2007
Available online 31 July 2007

Abstract

Flavonoids detected from a model legume plant, *Lotus japonicus* accessions Miyakojima MG-20 and Gifu B-129, were profiled using liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry (LC-FTICR/MS). Five flavonols and two anthocyanidins were detected as aglycones. LC-FTICR/MS facilitated simultaneous detection of 61 flavonoids including compounds that have not been reported previously. Chemical information of the peaks such as retention time, λ_{max}, m/z value of the quasi-molecular ion, m/z value of MS/MS fragment ions, and relative intensity of MS/MS fragments was obtained, along with the molecular formulas and conjugate structures. Fourteen were completely identified by comparison with authentic compounds. The high accuracy of m/z values, being 0.081 ppm between observed and theoretical values, allowed prediction of molecular formulas of unknown compounds with the help of isotope peak information for determination of chemical composition. Based on a predicted elemental composition, the presence of a novel nitrogen-containing flavonoid was proposed. A comparison of flavonoid profiles in flowers, stems, and leaves demonstrated that the flowers yielded the most complex profile, containing 30 flower-specific flavonoids including gossypetin glycosides and isorhamnetin glycosides. A comparison of flavonoid profiles between MG-20 and B-129 grown under the same conditions revealed that the accumulation of anthocyanins was higher in B-129 than MG-20, particularly in the stem. Developmental changes in the flavonoid profiles demonstrated that kaempferol glycosides increased promptly after germination. In contrast, quercetin glycosides, predominant flavonoids in the seeds, were not detectable in growing leaves.

Keywords: *Lotus japonicus*; Leguminoseae; Metabolite profiling; Flavonoids; LC-FTICR/MS; FT-MS

1. Introduction

Lotus japonicus has served as a model plant of the legumes along with *Medicago truncatula*. Genome sequencing of *L. japonicus* is nearly completed, and a large number of expressed sequence tags (ESTs) have been accumulated from various tissues. Among diverse accessions, Miyakojima MG-20 has been subjected to whole genome sequenc-

* Corresponding authors. Tel.: +81 43 290 2906; fax: +81 43 290 2905 (K. Saito), tel.: +81 438 52 3947; fax: +81 438 52 3948 (K. Aoki).
E-mail addresses: kaoki@kazusa.or.jp (K. Aoki), ksaito@faculty.chiba-u.jp (K. Saito).
[1] Present address: Kazusa DNA Research Institute, Kazusa-Kamatari 2-6-7, Kisarazu 292-0818, Japan.

0031-9422/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2007.06.017

ing (Sato et al., 2001; Nakamura et al., 2002; Asamizu et al., 2003; Kaneko et al., 2003; Kato et al., 2003). Substantial numbers of ESTs from roots and nodules have been accumulated from another accession, Gifu B-129 (Asamizu et al., 2004; Kouchi et al., 2004). A genetic linkage map was generated using these two accessions by PCR-based DNA markers (Asamizu et al., 2003). Furthermore, phylogenetic analysis revealed that the genetic distance among 16 *L. japonicus* accessions was largest between MG-20 and B-129 (Kawaguchi et al., 2001). The genomic information collected for MG-20 and B-129 provides a foundation to elucidate the mechanisms regulating biological functions and genetic diversity in *L. japonicus*.

Flavonoids are responsible for colors of plant organs, protective mechanisms against biotic and abiotic stresses, and regulation of normal plant development (Lazar and Goodman, 2006). A profiling of flavonoids and isoflavonoids in *M. truncatula* root and root-derived cell suspension culture has been reported recently (Farag et al., 2007). However, a comprehensive profiling of *L. japonicus* flavonoids has not been reported so far. Compounds belonging to the flavonoids are synthesized through the condensation reaction of three malonyl-CoAs and one *p*-coumaroyl-CoA, and then modified by dehydrogenation, hydroxylation, glycosylation, methylation, and acylation to yield various forms of flavonoids accumulated in plants. This multi-step biosynthetic process confers enormous structural diversity (Springob et al., 2003; Martens and Mithofer, 2005). From *L. japonicus*, genes for key enzymes of the flavonoid biosynthetic pathway, including chalcone isomerase (Shimada et al., 2003) and dihydroflavonol 4-reductase (Shimada et al., 2005), have been identified. Despite progress in identifying flavonoid synthesis-associated genes, information on metabolite profiles is still insufficient to depict the whole flavonoid pathway in this model legume plant.

A detailed analysis of flavonoid composition is required to elucidate complex flavonoid biosynthesis processes and genes involved in each process. To this end, a comprehensive flavonoid profiling appears to be a promising approach. Recently, Fourier transform ion cyclotron resonance mass spectrometry (FTICR/MS) has shown potential as a powerful technology for metabolite analysis (Aharoni et al., 2002; Hirai et al., 2004; Murch et al., 2004; Hirai et al., 2005; Tohge et al., 2005; Oikawa et al., 2006). The high mass-resolution ability of FTICR/MS facilitates the effective separation of plant metabolites even by direct infusion, though isomers with identical masses often seen in flavonoids could not be discriminated (Hall, 2006).

In the present study, we linked a liquid chromatography (LC) separation prior to FTICR/MS to overcome the limitation of isomer separation, and performed a comprehensive profiling of flavonoids from several developmentally different tissues of *L. japonicus* MG-20 and B-129. Accuracy of *m/z* value measurement by LC-FTICR/MS, combined with UV/visible absorption spectra, MS/MS fragmentation analysis, and isotopic ion analysis, allowed us to predict molecular formulas and structures with a high reliability.

2. Results

2.1. Optimization of HPLC conditions

We first optimized an elution protocol for the HPLC portion of LC-FTICR/MS. The elution protocol was evaluated by the separation of known flavonoids; kaempferol 3,7-di-*O*-rhamnoside (Kae-di-Rha), kaempferol

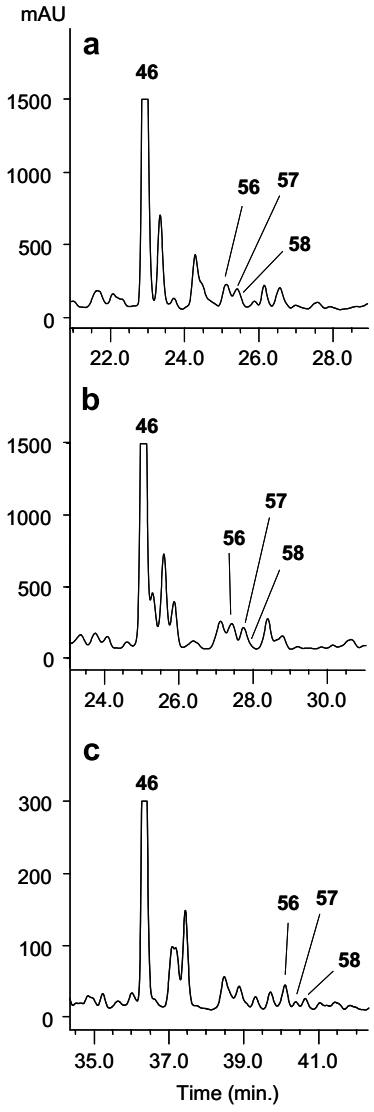


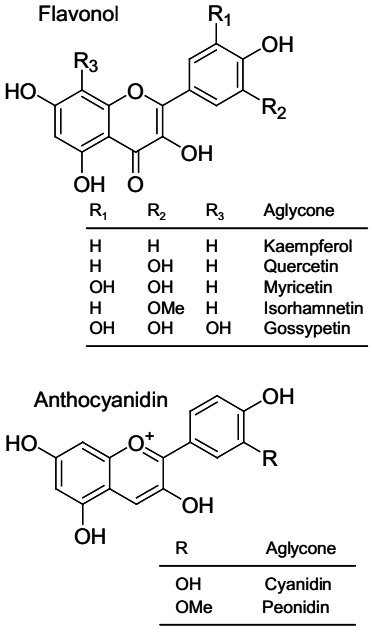
Fig. 1. Comparison of HPLC elution conditions for separating flavonoid metabolites. HPLC chromatograms of a stem extract of *Lotus japonicus* under elution conditions (a) A, (b) B, and (c) C, as described in Section 4. All chromatograms were monitored at 280 nm. The numbered peaks indicate **46**: kaempferol 3,7-di-*O*-rhamnoside (Kae-di-Rha), **56**: kaempferol 3-*O*-glucoside (Kae-Glc), **57**: isorhamnetin 3-*O*-glucoside (Irh-Glc), and **58**: quercetin 3-*O*-rhamnoside (Que-Rha). Elution condition C was used for the following analyses.

3-*O*-glucoside (Kae-Glc), isorhamnetin 3-*O*-glucoside (Irh-Glc), and quercetin 3-*O*-rhamnoside (Que-Rha) (Fig. 1). We compared the resolution of three elution protocols. First, using a conventional HPLC protocol for flavonoid profiling (Tohge et al., 2005), Kae-di-Rha and Kae-Glc were separated satisfactorily. However, Irh-Glc and Que-Rha eluted as a single peak (Fig. 1a). Secondly, we changed the column from ODS-80TM (4.6 × 150 mm) to ODS-100 V (4.6 × 250 mm), which had been packed with a more hydrophilic resin than that of ODS-80TM. However, Irh-Glc and Que-Rha still eluted as a single peak (Fig. 1b). Thirdly, we changed the mobile phase modifier from trifluoroacetic acid to formic acid. At the same time, the gradient program was modified (condition C in Section 4.3). By using this setting, Irh-Glc and Que-Rha were separated into two distinct peaks. Separation of the peaks corresponding to Kae-di-Rha and Kae-Glc was also improved (Fig. 1c). The third protocol was thus used in the following analyses.

2.2. Detailed analysis of mass spectral data

To determine flavonoid aglycones, crude extracts from tissues were acid hydrolyzed and then analyzed by HPLC. Five flavonol aglycones (kaempferol, quercetin, isorhamnetin, myricetin, and gossypetin) and two anthocyanidins (cyanidin and peonidin) were detected (Fig. 2).

LC-FTICR/MS was then used for analysis of flavonoids. The extracts were analyzed in positive-ion electrospray ionization (ESI) mode, as positive-ion mode yielded more fragmentation information than negative-ion mode as reported in Farag et al. (2007). As shown in HPLC chromatograms (Figs. 3 and 4), 61 peaks were detected collectively from mature plants (90-day-old plants) of MG-20 and B-129. Peak information including retention time, λ_{max} of the UV/visible spectrum, and the *m*/*z* value of the observed ions is summarized in Table 1. The identities of 14 peaks were confirmed by comparing the experimental values with those of the authentic compounds (underlined peaks in Table 1). The accuracy of *m*/*z* measurement by FTICR/MS was Δppm 0.2–1.4 even without calibration using internal standards. By using the data correction method based on that reported by Oikawa et al. (2006), the average Δppm for the *m*/z values of these 14 metabolites was improved to 0.118. Among these 14 peaks, four pairs of peaks, **5** and **7**, **15** and **18**, **48** and **49**, and **56** and **58**, had identical *m*/*z* values (Table 1). However, all these pairs were separated by HPLC. This result clearly demonstrated the advantage of linking LC prior to MS for the comprehensive analysis of flavonoids.



R_1	R_2	R_3	Aglycone
H	H	H	Kaempferol
H	OH	H	Quercetin
OH	OH	H	Myricetin
H	OMe	H	Isorhamnetin
OH	OH	OH	Gossypetin

R	Aglycone
OH	Cyanidin
OMe	Peonidin

Fig. 2. Structures of flavonoid aglycones found in aerial parts of *Lotus japonicus*.

For 47 other peaks for which authentic compounds were not available, we attempted to provide annotations of the putative molecular formulas and conjugate structures (Table 1). The average Δppm of the observed *m*/*z* against the theoretical *m*/*z* of the best-hit molecular formula was 0.081. Conjugate structures were assigned based on the detected aglycones and MS/MS fragmentation profiles. Aglycones were assigned according to the smallest *m*/*z* of the MS/MS fragments and λ_{max} of the UV absorption spectrum. For example, both peak **1** and peak **3** had smallest MS/MS fragments of *m*/*z* 287.1. However, λ_{max} of peak **1** was 348 nm, while that of peak **3** was 516 nm, demonstrating that the aglycones of peak **1** and peak **3** were kaempferol and cyanidin, respectively. The conjugate moieties were determined by *m*/*z* differences between MS/MS fragments.

We also included relative intensity of MS/MS fragment ions in the list of peak information (Table 1). We recognized that three pairs of multi-glycoside peaks (**1** and **2**, **22** and **23**, and **39** and **42**) had identical *m*/*z* values of precursors and MS/MS ions. However, relative intensity patterns were significantly different. The difference was reproducible in six measurements.

2.3. Gossypetin glycosides

We detected gossypetin glycosides, corresponding to peaks **21** and **26**, in flowers of MG-20 and B-129 (Table 1). MS/MS fragmentation patterns demonstrated that peaks **21** and **26** were hexose-conjugated gossypetin. We purified gossypetin glycosides LJ-YP1 and LJ-YP2 from peak **21** and **26**, respectively. UV absorption spectra of acid-hydrolyzed products of the purified compounds were identical to that of authentic gossypetin (Fig. 5). The NMR spectroscopic data indicated that the structural core of LJ-YP1 was a polyhydroxygenated flavonoid. The characteristic four aromatic protons at δ 7.74 (dd, H-6′), 7.64 (d, H-2′), 6.82 (d, H-5′), and 6.28 (s, H-6) as well as the chemical shifts of fifteen sp^2 carbons ascribable to flavonoid part established it to be gossypetin (3,5,7,8,3′,4′-hexahydroxyflavone). Further, the ^{1}H and ^{13}C NMR

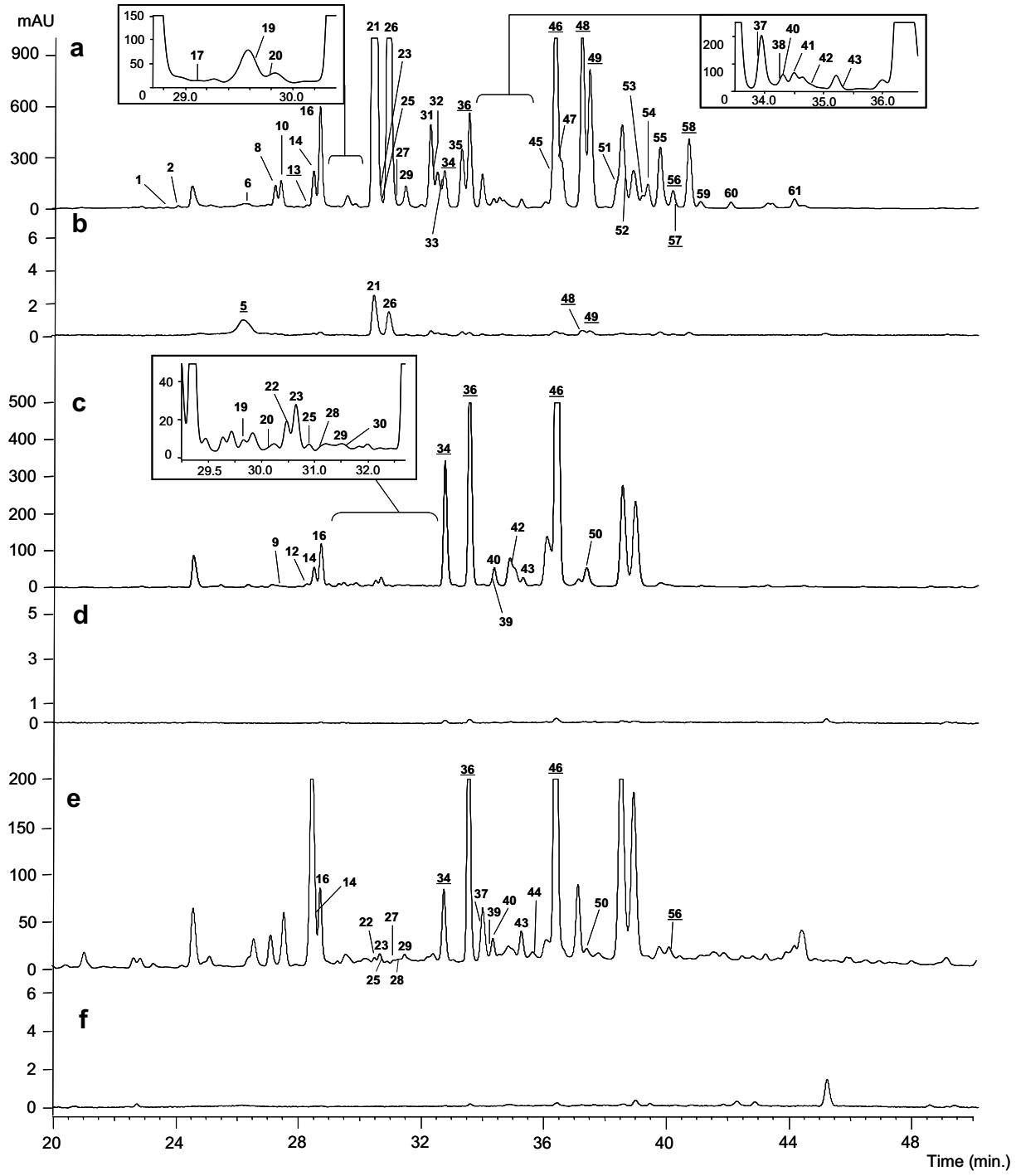


Fig. 3. HPLC elution profiles of flavonoids from *Lotus japonicus* accession Miyakojima MG-20. (a) flower, 280 nm, (b) flower, 520 nm, (c) leaf, 280 nm (d) leaf, 520 nm, (e) stem, 280 nm, (f) stem, 520 nm. The peak numbers correspond to flavonoids shown in Table 1.

spectroscopic data showed that this compound possessed one sugar, the structure of which was identical to galactose by comparison of the ^{13}C NMR data. The HMBC correlation between H-1″ (δ 5.36) in galactose and C-3 (δ 133.8) in the gossypetin moiety established that the sugar was attached to the C-3 position of the flavonol. The coupling constant ($J = 7.6$ Hz) of the anomeric proton showed the β-anomer form of galactose. Therefore, the structure of LJ-YP1 was assigned as gossypetin 3-*O*-β-galactoside.

The spectroscopic data as well as chromatographic behavior of LJ-YP2 were very similar to those of LJ-YP1. It showed four aromatic protons at δ 7.68 (overlapped, H-2′), 7.67 (overlapped, H-6′), 6.85 (d, H-5′), and 6.28 (s, H-6) and 15 sp^2 carbons ascribable to gossypetin. The sugar component was identical to glucose by comparison of the ^{13}C NMR spectroscopic data. The HMBC correlation between H-1″ (δ 5.45) in glucose and C-3 (δ 133.7) in gossypetin established that the sugar was attached to C-3

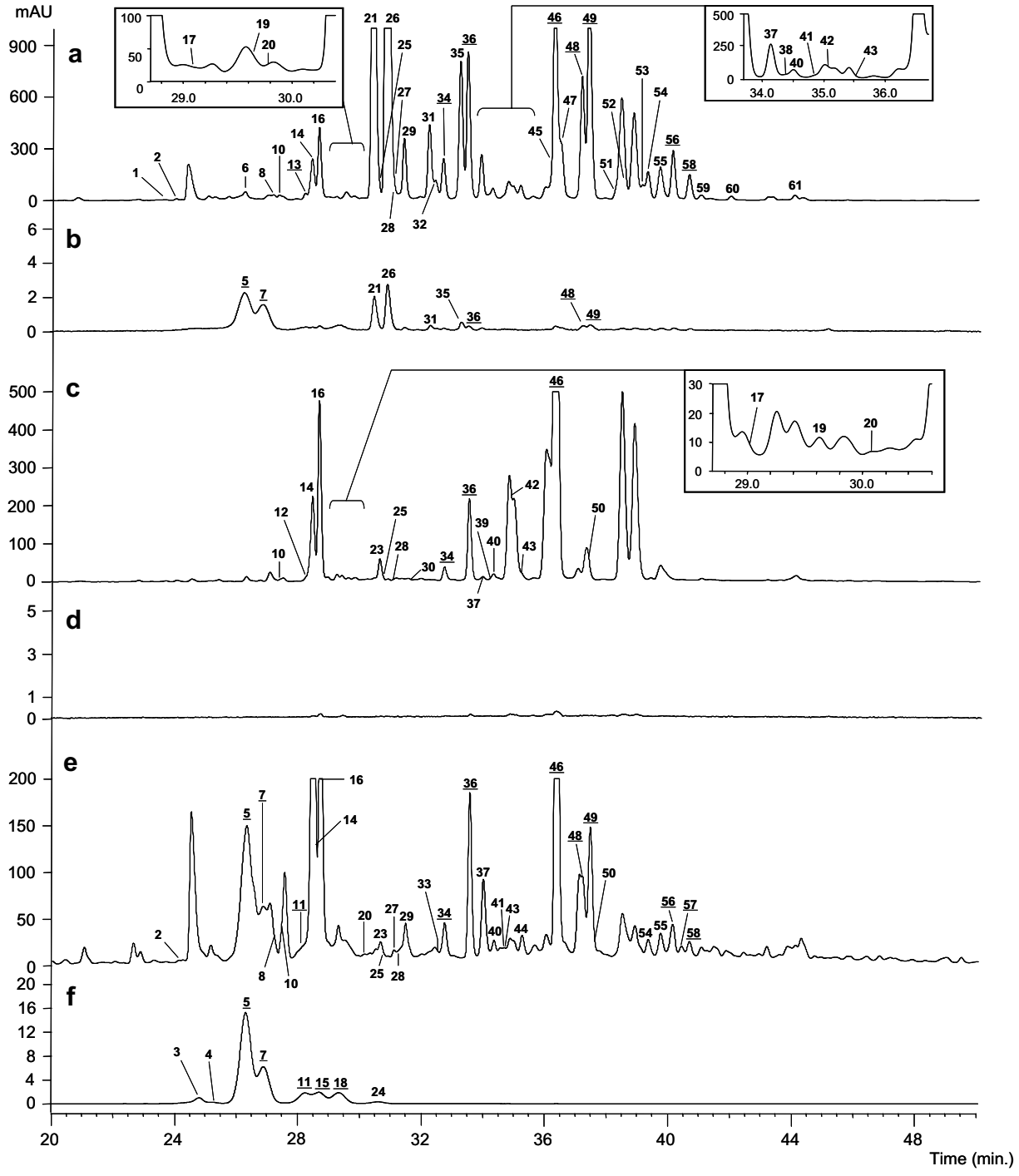


Fig. 4. HPLC elution profiles of flavonoids from *Lotus japonicus* accession Gifu B-129. (a) flower, 280 nm, (b) flower, 520 nm, (c) leaf, 280 nm (d) leaf, 520 nm, (e) stem, 280 nm, (f) stem, 520 nm. The peak numbers correspond to flavonoids shown in Table 1.

of the flavonol. The coupling constant ($J = 7.3$ Hz) of the anomeric proton showed the β-anomer form of glucose. Therefore the structure of LJ-YP2 was assigned as gossypetin 3-*O*-β-glucoside.

Although the presence of gossypetin 3-*O*-galactoside and gossypetin 3-*O*-glucoside in higher plants has been reported previously (Seshadri and Thakur, 1961; Harborne, 1969), full-assignment of the NMR spectra has not been reported; these are reported in Section 4.

2.4. Isorhamnetin glycoside and a putative nitrogen-containing flavonoid

We confirmed the presence of isorhamnetin 3-*O*-glucoside (peak **57**) using authentic standard (Table 1). Although the presence of isorhamnetin has been reported in the literature (Jay et al., 1978), a glycoside based on this aglycone has not been proven (Reynaud and Lussignol, 2005). Our result provides experimental evidence for the

Table 1
Annotation for flavonoid peaks detected from *Lotus japonicus*

Peak No.	Organs		Rt (min.)	λ_{max} (nm)	$[M+H]^+$	Δppm	$[MS/MS]^+$ (relative intensity, %)	Molecular formula
	MG-20	B-129						
1	F	F	23.77	266, 348	773.213393	−0.120	611.0 (100), 449.0 (10.5), 287.1 (43.6)	$C_{33}H_{40}O_{21}$ (Kae-Hex, -Hex, -Hex)
2	F, S	F, S	24.10	266, 348	773.213535	0.064	610.9 (100), 449.0 (37.8), 287.1 (89.7)	$C_{33}H_{40}O_{21}$ (Kae-Hex, -Hex, -Hex)
3		S	24.76	280, 516	611.160624	−0.062	287.1 (100)	$C_{27}H_{30}O_{16}$ (Cya-Hex, -Hex)
4		S	25.16	274, 516	611.160675	0.021	287.1 (100)	$C_{27}H_{30}O_{16}$ (Cya-Hex, -Hex)
5[a]	F	F, S	26.27	280, 518	449.107776	−0.139	287.1 (100)	Cyanidin 3-*O*-galactoside
6	F	F	26.33	256, 350	627.155567	−0.015	465.0 (100), 303.1 (26.7)	$C_{27}H_{30}O_{17}$ (Que-Hex, -Hex)
7[a]		F, S	26.85	280, 518	449.107821	−0.039	287.1 (100)	Cyanidin 3-*O*-glucoside
8	F	F, S	27.24	270, 352	773.213491	0.007	610.8 (28.0), 449.0 (100), 303.1 (78.0)	$C_{33}H_{40}O_{21}$ (Que-6DH, -Hex, -Hex)
9	L		27.30	266, 346	611.160606	−0.092	449.1 (100), 287.1 (29.7)	$C_{27}H_{30}O_{16}$ (Kae-Hex, -Hex)
10	F	F, L, S	27.42	256, 354	773.213440	−0.059	610.9 (30.4), 449.0 (100), 303.0 (79.2)	$C_{33}H_{40}O_{21}$ (Que-6DH, -Hex, -Hex)
11[a]		S	28.18	280, 518	419.097254	−0.047	287.1 (100)	Cyanidin 3-*O*-arabinoside
12	L	L	28.22	266, 344	943.282682	0.057	511.2 (100), 433.1 (5.8), 349.1 (13.7), 331.1 (7.3), 313.1 (7.7), 287.1 (9.5)	$C_{40}H_{50}O_{24}N_2$ (Kae - 6DH, -Hex -Un)
13[a]	F	F,	28.28	268, 322	611.160633	−0.047	449.1 (100), 287.1 (23.3)	Kaempferol 3,7-di-*O*-glucoside
14	F, L, S	F, L, S	28.50	266, 348	757.218649	0.103	611.0 (4.6), 595.0 (20.6), 449.0 (5.6), 433.0 (100), 287.2 (86.7)	$C_{33}H_{40}O_{20}$ (Kae-6DH, -Hex-Hex)
15[a]		S	28.65	280, 326, 518	463.123466	−0.049	301.1 (100)	Peonidin 3-*O*-galactoside
16	F, L, S	F, L, S	28.71	266, 348	757.218519	−0.069	610.9 (0.9), 594.9 (26.8), 449.1 (1.7), 433.0 (100), 287.1 (79.1)	$C_{33}H_{40}O_{20}$ (Kae-6DH, -Hex-Hex)
17	F	F, L	29.02	258, 348	743.202939	0.024	610.9 (45.2), 465.1 (2.4), 449.0 (100), 303.2 (65.3)	$C_{32}H_{38}O_{20}$ (Que-6DH, -Hex)
18[a]		S	29.27	268, 516	463.123452	−0.079	301.1 (100)	Peonidin 3-*O*-glucoside
19	F, L	F, L	29.67	266, 348	829.239716	0.019	433.1 (82.0), 287.1 (100)	$C_{36}H_{44}O_{22}$ (Kae-6DH, -Un)
20	F, L	F, L, S	30.10	268, 344	843.255424	0.087	680.9 (100), 433.1 (39.9), 287.1 (40.9)	$C_{37}H_{46}O_{22}$ (Kae-6DH, -Hex, -Hex-Malonyl)
21	F	F	30.42	274, 342	481.097538	−0.270	319.1 (100)	$C_{21}H_{20}O_{13}$ (Gossy-Hex)
22	L, S		30.49	266, 348	727.208058	0.071	595.1 (28.5), 581.2 (6.7), 449.1 (7.4), 433.0 (100), 287.1 (48.1)	$C_{32}H_{38}O_{19}$ (Kae-6DH, -Hex-Pen)
23	F, L, S	L, S	30.61	266, 348	727.208020	0.019	594.9 (42.3), 433.1 (100), 287.2 (45.1)	$C_{32}H_{38}O_{19}$ (Kae-6DH, -Hex-Pen)
24		S	30.66	298, 518	433.112866	−0.133	301.1 (100)	$C_{21}H_{20}O_{10}$ (Peo-Pen)
25	F, L, S	F, L, S	30.72	266, 346	843.255263	−0.104	594.9 (3.3), 433.0 (100), 287.1 (93.6)	$C_{37}H_{46}O_{22}$ (Kae-6DH, -Hex, -Hex-Malonyl)
26	F	F	30.84	274, 340	481.097539	−0.268	319.1 (100)	$C_{21}H_{20}O_{13}$ (Gossy-Hex)
27	F, S	F, S	31.08	268, 354	611.160722	0.098	465.0 (4.5), 449.0 (100), 303.1 (29.8)	$C_{27}H_{30}O_{16}$ (Que-6DH, -Hex)
28	F, L, S	F, L, S	31.15	264, 344	843.255336	−0.017	594.8 (2.7), 433.1 (100), 287.1 (90.1)	$C_{37}H_{46}O_{22}$ (Kae-6DH, -Hex, -Hex-Malonyl)
29	F, L, S	F, S	31.46	256, 354	611.160685	0.038	465.0 (2.4), 449.0 (100), 303.1 (30.8)	$C_{27}H_{30}O_{16}$ (Que-6DH, -Hex)

30	L	L	31.60	266, 340	799.229202	0.083	595.1 (5.3), 433.0 (94.7), 287.1 (100)	$C_{35}H_{42}O_{21}$ (Kae-6DH, -Hex, -Un)
31	F	F	32.22	224, 274, 300, 326	465.102695	−0.125	303.2 (100)	$C_{21}H_{20}O_{12}$ (Unidentified)
32	F	F	32.44	276, 338	451.087079	−0.057	319.1 (100)	$C_{20}H_{18}O_{12}$ (Gossy-Pen)
33	F	S	32.63	258, 354	627.155540	−0.058	465.0 (14.4), 303.1 (100)	$C_{27}H_{30}O_{17}$ (Que-Hex, -Hex)
34[a]	F, L, S	F, L, S	32.76	266, 346	595.165660	−0.147	449.1 (3.6), 433.0 (100), 287.2 (28.3)	Kaempferol 3-*O*-galactosyl-7-*O*-rhamnoside
35	F	F	33.32	224, 276, 300, 328	465.102713	−0.086	303.1 (100)	$C_{21}H_{20}O_{12}$ (Unidentified)
36[a]	F, L, S	F, L, S	33.53	266, 346	595.165566	−0.305	449.0 (1.5), 433.0 (100), 287.1 (31.1)	Kaempferol 3-*O*-glucosyl-7-*O*-rhamnoside
37	F, S	F, L, S	33.98	256, 350	595.165671	−0.128	449.0 (100), 303.2 (8.6)	$C_{27}H_{30}O_{15}$ (Que-6DH, -6DH)
38	F	F	34.16	264, 356	481.097657	−0.022	319.1 (100)	$C_{21}H_{20}O_{13}$ (Myr-Hex)
39	L, S	L	34.23	242, 322	741.223689	0.044	595.0 (71.3), 579.1 (21.5), 433.0 (93.6), 287.1 (100)	$C_{33}H_{40}O_{19}$ (Kae-6DH, -6DH, -Hex)
40	F, L, S	F, L, S	34.33	266, 344	565.155198	0.027	433.0 (100), 419.1 (7.6), 287.1 (12.5)	$C_{26}H_{28}O_{14}$ (Kae-6DH, -Pen)
41	F	F, S	34.66	266, 348	611.160691	0.047	449.0 (12.9), 287.1 (100)	$C_{27}H_{30}O_{16}$ (Kae-Hex, -Hex)
42	F, L	F, L	34.80	266, 338	741.223687	0.041	595.1 (14.5), 579.0 (19.9), 433.1 (100), 287 (81.8)	$C_{33}H_{40}O_{19}$ (Kae-6DH, -6DH, -Hex)
43	L, S	F, L, S	35.27	264, 340	681.166151	0.014	535.0 (16.4), 433.0 (77.2), 287.1 (100)	$C_{30}H_{32}O_{18}$ (Kae-6DH, -Hex-Malonyl)
44	S	S	35.67	244, 338	711.176714	0.011	565.0 (14.9), 463.1 (95.3), 317.1 (100)	$C_{31}H_{34}O_{19}$ (Irh-6DH, -Hex-Malonyl)
45	F	F	36.20	272, 362	495.113248	−0.141	333.1 (100)	$C_{22}H_{22}O_{13}$ (Unidentified)
46[a]	F, L, S	F, L, S	36.32	264, 342	579.170581	−0.435	433.0 (100), 287.1 (7.4)	Kaempferol 3,7-di-*O*-rhamnoside
47	F	F	36.46	270, 358	495.113255	−0.127	333.1 (100)	$C_{22}H_{22}O_{13}$ (Unidentified)
48[a]	F	F, S	37.16	256, 354	465.102696	−0.123	303.1 (100)	Quercetin 3-*O*-galactoside
49[a]	F	F, S	37.43	256, 354	465.102699	−0.116	303.1 (100)	Quercetin 3-*O*-glucoside
50	L, S	L, S	37.40	268, 332	841.276037	−0.058	433.1 (96.0), 287.0 (100)	$C_{38}H_{40}O_{21}$ (Kae-6DH, -Un)
51	F	F	38.30	270, 362	479.118347	−0.117	317.1 (100)	$C_{22}H_{22}O_{12}$ (Un-Hex)
52	F	F	38.77	270, 358	465.102742	−0.024	333.1 (100)	$C_{21}H_{20}O_{12}$ (Unidentified)
53	F	F	39.10	266, 334	479.118353	−0.105	317.1 (100)	$C_{22}H_{22}O_{12}$ (Un-Hex)
54	F	F, S	39.34	266, 344	449.107807	−0.070	287.1 (100)	$C_{21}H_{20}O_{11}$ (Kae-Hex)
55	F	F, S	39.74	256, 352	435.092191	0.006	303.1 (100)	$C_{20}H_{18}O_{11}$ (Que-Pen)
56[a]	F, S	F, S	40.11	266, 348	449.107847	0.019	287.1 (100)	Kaempferol 3-*O*-glucoside
57[a]	F	S	40.43	256, 352	479.118352	−0.107	317.1 (100)	Isorhamnetin 3-*O*-glucoside
58[a]	F	F, S	40.68	256, 348	449.107834	−0.010	303.2 (100)	Quercetin 3-*O*-rhamnoside
59	F	F	41.12	266, 348	449.107828	−0.023	317.1 (100)	$C_{21}H_{20}O_{11}$ (Un -Pen)
60	F	F	42.00	266, 344	419.097272	−0.004	287.1 (100)	$C_{20}H_{18}O_{10}$ (Kae-Pen)
61	F	F	44.06	264, 342	433.112910	−0.032	287.2 (100)	$C_{21}H_{20}O_{10}$ (Kae-6DH)

F; Flower, L; Leaf, S; Stem, Rt; retention time, Kae; kaempferol, Que; quercetin, Irh; isorhamnetin, Myr; Myricetin, Gossy; gossypetin, Hex; hexose, 6DH; 6-deoxyhexose, Pen; pentose, Un; unidentified moiety.

[a] Metabolite identical to authentic compound.

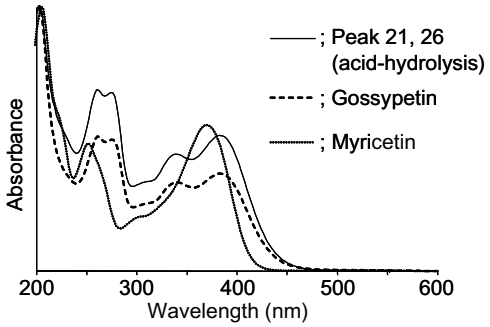
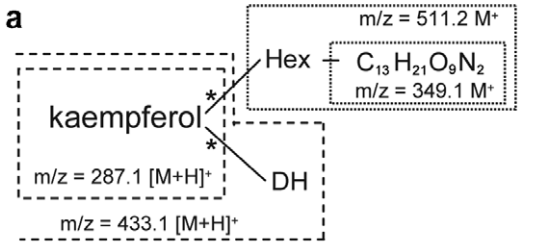


Fig. 5. UV absorption spectra of acid-hydrolyzed products from metabolites in peaks **21** and **26**. The spectra are identical to that of authentic gossypetin (broken curve), and different from that of myricetin (dotted curve), indicating that the metabolites in these peaks had the gossypetin aglycone.

presence of isorhamnetin glycosides in *L. japonicus*. The aglycone of the metabolite in peak **44** was also confirmed to be isorhamnetin by MS^3 pattern (data not shown).

The observed *m/z* of peak **12** matched the theoretical *m/z* of the predicted molecular formula $C_{40}H_{51}O_{24}N_2$ (Δppm 0.058). The proposed molecular formula did not have any significant "hits" in the Dictionary of Natural Products (DNP). Fig. 6a illustrates the putative structure of the metabolite deduced from its MS/MS fragmentation pattern. Two MS/MS fragments, *m/z* 331.1 and *m/z* 313.1, were probably generated by dehydration of the *m/z* 349.1 fragment, which likely had a nitrogen-containing group. The relative intensities of isotopic peaks supported the proposed molecular formula (Yergey, 1983; Kind and Fiehn, 2006). The observed *m/z* and relative intensity of the $[(M+1)+H]^+$ peak associated with peak **12** matched the theoretical *m/z* (Δppm = 0.29) and relative intensity (2.5% difference; Fig. 6b). The theoretical relative intensity of the $[(M+1)+H]^+$ peak for other candidate formulas, $C_{52}H_{47}O_{17}$ (best-hit formula for elemental composition C, H, O), or $C_{47}H_{64}O_6P_5S_2$ (best-hit formula for elemental composition C, H, O, P, N, S) did not match the observed value (Fig. 6b). Collectively, these data indicated that $C_{40}H_{51}O_{24}N_2$ matched the experimental data of peak **12**, although it was not proven.



b

Observed [m/z (rel. Int.)]	Theoretical [m/z (rel. Int.)]	Molecular Formula
943.282682 (100)	943.282627 (100)	$[(C_{40}H_{50}O_{24}N_2)+H]^+$
944.286258 (41.84±2.0)	944.285981 (43.3)	$[(C_{39}{}^{13}C_1H_{50}O_{24}N_2)+H]^+$
	943.280776 (100)	$[(C_{52}H_{47}O_{17})+H]^+$
	944.284131 (56.2)	$[(C_{51}{}^{13}C_1H_{47}O_{17})+H]^+$
	943.282690 (100)	$[(C_{47}H_{64}O_6P_5S_2)+H]^+$
	944.286045 (50.8)	$[(C_{46}{}^{13}C_1H_{64}O_6P_5S_2)+H]^+$

Fig. 6. (a) Putative structure of the metabolite in peak **12**, and (b) comparison of relative intensity of $[(M+1)+H]^+$ peaks. In (a), the smallest fragment, *m/z* 287, demonstrated kaempferol aglycone. MS/MS fragments, *m/z* 313.1 and 331.1, were probably derived from *m/z* 349.1 by dehydration. Hex; hexose, DH; 6-deoxyhexose. In (b), observed isotope peaks (left) and theoretically simulated isotope peaks (right) were indicated for hypothesized molecular formula $C_{40}H_{51}O_{24}N_2$ and two other proposed formulas. Numbers in parentheses represent relative intensity (mean ± SD) of the $[(M+1)+H]^+$ peak to the $[M+H]^+$ peak.

2.5. Unidentified peaks

Despite detailed measurement using LC-FTICR/MS, we failed to annotate twelve peaks (Table 1). Peaks **31**, **35**, **45**, **47**, **51**, **52**, **53**, and **59** appeared to have unknown aglycones. For example, peaks **31** and **35** had aglycones of *m/z* 303; however, UV/visible absorption spectra were different from that of quercetin (Fig. 7). Peaks **12**, **19**, **30**, and **50** appeared to have unknown conjugated moieties, although their aglycones were identified as kaempferol. The proposed molecular formula for peak **30** ($C_{35}H_{42}O_{21}$) matched astragalin, populnin, and trifolin, implying that the unidentified conjugated moiety could contain two hexoses and one acetyl group.

We checked whether flavonoid and isoflavonoid metabolites reported in *M. truncatula* (Farag et al., 2007) were present in *L. japonicus* or not. We did not detect aglycones of isoflavonoids, liquiritigenin, isoliquiritigenin, or naringenin in the HPLC data of the acid hydrolyzed products. We indeed detected ions with identical *m/z* values with the reported metabolites. However, their intensities were too weak to identify. Collectively, we did not confirm the pres-

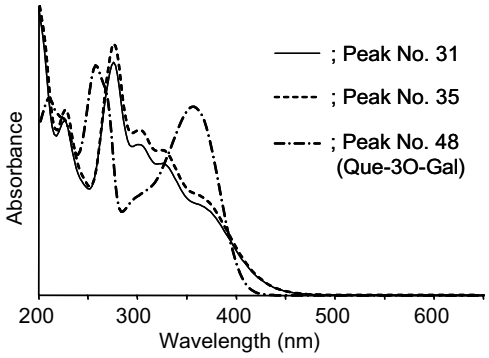


Fig. 7. UV absorption spectra of metabolites in peaks **31** and **35**. These spectra were different from that of quercetin (chain curve), although the MS/MS fragmentation profile demonstrated that peaks **31** and **35** had an aglycone of *m/z* 303 that was identical to the *m/z* value of quercetin. The aglycone of metabolites in peaks **31** and **35** remains unidentified.

ence of flavonoids and isoflavonoids reported in *M. truncatula* by Farag et al. (2007) in the aerial part of *L. japonicus*, although we could not exclude the possibility that trace amounts of those metabolites were present.

2.6. Flavonoid profiles in MG-20 and B-129

Differences in pigmentation between MG-20 and B-129 have been reported (Kawaguchi et al., 2001). Based on the analysis of mass spectral data, we compared flavonoid profiles of MG-20 and B-129 in detail. In MG-20 (Fig. 3), flavonoid composition was most complicated in flowers. The elution profile of A_{280} peaks showed that MG-20 flowers contained 46 flavonoids (Fig. 3a), and 30 were flower-specific. Hexose-conjugated gossypetin (peak **21** and **26**) and hexose-conjugated quercetin (peak **48** and **49**) were the main flavonoids in MG-20 flowers. One of the A_{520} peaks (peak **5**, Fig. 3b) was confirmed to be cyanidin 3-*O*-galactoside. The main flavonoid components in MG-20 leaves (Fig. 3c) were kaempferol derivatives, including 4 leaf-specific peaks. MG-20 leaves did not contain a detectable amount of anthocyanin (Fig. 3d). In stems of MG-20 (Fig. 3e), kaempferol derivatives were the main flavonoid components as in leaves. No anthocyanin was detected in MG-20 stems (Fig. 3f).

Table 2
Flavonoid content in seeds of MG-20 and B-129

Metabolite	MG-20 (ng/mg fresh weight)	B-129 (ng/mg fresh weight)
Kaempferol 3-*O*-glucosyl-7-*O*-rhamnoside	19.6 ± 3.2	n.d.
Kaempferol 3,7-di-*O*-rhamnoside	58.0 ± 8.0	16.7 ± 1.6
Quercetin 3-*O*-galactoside	177.9 ± 18.4	46.5 ± 3.3
Quercetin 3-*O*-glucoside	51.3 ± 4.2	86.2 ± 9.0
Quercetin 3-*O*-rhamnoside	1176.6 ± 87.5	1090.9 ± 65.5

n.d., not detected.
Values represent means ±SE of three repeats.

From B-129 flowers, 17 flower-specific flavonoids were detected (Fig. 4a). The elution profile of A_{520} peaks (Fig. 4b) showed that B-129 flowers contained cyanidin 3-*O*-galactoside (peak **5**) and cyanidin 3-*O*-glucoside (peak **7**). As in MG-20 flowers, hexose-conjugated gossypetin (peaks **21** and **26**) and hexose-conjugated quercetin (peaks **48** and **49**) were the main flavonoids in leaves, B-129 contained 21 flavonoids (Fig. 4c) and no detectable anthocyanins (Fig. 4d). In stems, 37 flavonoids (Fig. 4e) including 7 anthocyanins (Fig. 4f) were detected.

By these results, the distinct pigmentation of B-129 (Kawaguchi et al., 2001) was characterized at the flavonoid composition level. The most remarkable difference between MG-20 and B-129 was seen in the stems under our growth conditions. Eight anthocyanins (peaks **3**, **4**, **5**, **7**, **11**, **15**, **18**, and **24**) were detected in B129 stems, whereas no anthocyanin was detected in MG-20 stems. In B-129 flowers, two cyanidin glucosides (peaks **5** and **7**) were also detected.

2.7. Developmental changes in flavonoid profile

To monitor developmental changes in flavonoids, the flavonoid contents in dry seeds (Table 2) and leaves (Table 3) were analyzed quantitatively. In dry seeds, quercetin derivatives were much more abundant than kaempferol derivatives. In particular, quercetin 3-*O*-rhamnoside was most abundant (Table 2).

In leaves, kaempferol glycosides accumulated to a remarkable degree (Table 3). The accumulation level of kaempferol 3,7-di-*O*-rhamnoside was apparently higher than other kaempferol glycosides. Kaempferol 3-*O*-glucosyl-7-*O*-rhamnoside and kaempferol 3-*O*-galactosyl-7-*O*-rhamnoside showed nearly the same developmental profiles as kaempferol 3,7-di-*O*-rhamnoside, although the maximum accumulation levels were much lower. These three kaempferol glycosides accumulated promptly after germination as reported for total kaempferol in the primary leaf of *Glycine max* (Graham, 1991). In contrast, kaempferol 3-

Table 3
Flavonol contents in leaves of MG-20 and B-129

Day after sowing	7 (ng/mg fresh weight)	14 (ng/mg fresh weight)	21 (ng/mg fresh weight)	30 (ng/mg fresh weight)	60 (ng/mg fresh weight)	90 (ng/mg fresh weight)
Kaempferol 3-*O*-galactosyl-7-*O*-rhamnoside						
B-129 Leaf	12.0 ± 0.8	36.1 ± 11.4	79.1 ± 11.3	67.3 ± 4.8	77.9 ± 4.3	86.7 ± 13.9
MG-20 Leaf	24.1 ± 20.1	147.2 ± 36.0	234.9 ± 17.2	197.3 ± 46.5	416.0 ± 11.4	403.8 ± 109.0
Kaempferol 3-*O*-glucosyl-7-*O*-rhamnoside						
B-129 Leaf	32.0 ± 5.7	80.0 ± 18.0	217.0 ± 13.6	338.9 ± 26.1	381.7 ± 7.5	359.2 ± 37.4
MG-20 Leaf	25.8 ± 22.2	183.6 ± 19.3	352.7 ± 11.3	395.9 ± 80.2	803.3 ± 26.3	638.8 ± 172.5
Kaempferol 3,7-di-*O*-rhamnoside						
B-129 Leaf	393.0 ± 85.9	1050.0 ± 121.7	1059.6 ± 153.0	1521.8 ± 123.9	1870.6 ± 68.4	1508.6 ± 72.6
MG-20 Leaf	393.0 ± 337.2	1147.5 ± 52.5	1260.8 ± 110.1	1141.0 ± 216.9	1984.5 ± 81.3	1233.3 ± 256.5
Kaempferol 3-*O*-glucoside						
B-129 Leaf	n.d.	n.d.	n.d.	5.6 ± 0.7	4.9 ± 0.4	n.d.
MG-20 Leaf	n.d.	n.d.	n.d.	n.d.	2.9 ± 0.2	2.9 ± 0.4

n.d.: Not detected.
Values represent means ± SE of three repeats.

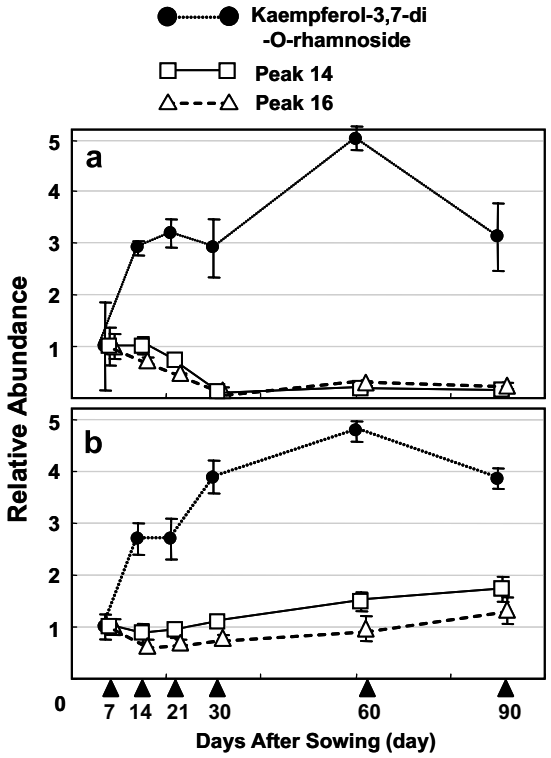


Fig. 8. Developmental change in relative content of kaempferol glycosides in leaves. (a) MG-20, (b) B-129. Relative abundance was calculated by dividing peak area by the peak area 7 days after sowing.

O-glucoside did not accumulate as much as the three other derivatives, demonstrating that kaempferol accumulated in a compound-selective manner. Other flavonoids, quercetin derivatives, isorhamnetin derivatives, and anthocyanins were not detectable in growing leaves (data not shown).

To further investigate the accumulation pattern of kaempferol derivatives, peaks **14** and **16** (both kaempferol conjugated by one deoxyhexose and two hexose moieties) were also analyzed. In both MG-20 and B-129, peaks **14** and **16** accumulated rapidly before day 7 after sowing (Fig. 8), demonstrating that these kaempferol glycosides also accumulated rapidly during germination. There was a slight difference between accumulation profiles of MG-20 and B-129. In MG-20, they decreased 14 days after sowing (Fig. 8a). In B-129, the level stayed constant after 14 days (Fig. 8b).

These results demonstrated that quercetin glycosides and kaempferol glycosides accumulated selectively in seeds and leaves, respectively. Among 12 quercetin glycosides detected in *L. japonicus*, only one species was detected in MG-20 leaves, and three in B-129 leaves (Table 1). This suggests that the biosynthesis of quercetin and kaempferol are regulated spatially in a tissue-specific manner.

3. Discussion

The results presented in this study illustrate that LC-FTICR/MS is a useful tool for predicting putative molecular formulas and structures. In molecular formula prediction, the high accuracy of mass measurement by LC-FTICR/MS effectively narrowed down the candidate formulas. The isotope peak profile turned out to be informative in molecular formula prediction, as the relative intensity of isotopic peaks could provide supporting information with respect to the molecular composition. The high mass resolution of LC-FTICR/MS enabled unambiguous assignment of isotopic peaks and comparison of their relative intensity. The relative intensity of MS/MS fragments was potentially informative in predicting molecular structure. It has been reported that the relative intensity of MS/MS fragments depended on linkage and branching structure of an oligosaccharide (Pfenninger et al., 2002; Zaia et al., 2003). Likewise, the differences in relative intensity patterns observed for flavonoids might reflect differences in the position and structure of the sugar linkage.

We performed integrative interpretation of the LC-FTICR/MS data to predict molecular formulas and structures of 61 flavonoids. Certainty of the prediction was classified into three classes. First, the molecular formulas and structures were confirmed by authentic compounds (14 flavonoids) and by experimentally determined structure using NMR (2 flavonoids). Second, molecular formulas and structures were supported by the combination of LC-FTICR/MS data, which matched consistently with the predictions (33 flavonoids). Third, molecular formulas and structures were unambiguously predicted, although aglycones or conjugate moieties remained to be identified (12 flavonoids). For the individual flavonoids in the second and the third classes, elucidation of the structures (e.g., position of the conjugate sugars) with other analytical methods is left for future study. Nevertheless, a prediction of molecular formulas and structures by integrative data interpretation was sufficiently reliable to identify flavonoid peaks across samples. It facilitated an efficient comparison of flavonoid compositions in different tissues (Table 1).

The comparative analysis of flavonoid profiles indicated that MG-20 and B-129 had distinct profiles in stems and young leaves, demonstrating the presence of metabolic diversity between two genetically distant accessions. The tissue-dependent accumulation pattern of flavonoids was also demonstrated by comparing flavonoid profiles in flowers, stems, leaves, and dry seeds. Particularly, dry seeds contained abundant quercetin glycosides, not much of kaempferol glycosides (Table 2). It has been reported that quercetin glycosides accumulate in the outer integument of Arabidopsis seed (Nesi et al., 2002; Pourcel et al., 2005). Similarly, in *L. japonicus* seed, quercetin derivatives are probably present in the testa.

3.1. Concluding remarks

In addition to identifying peaks across samples, a prediction of molecular formulas and structures will facilitate efficient mapping of the detected metabolites on flavonoid biosynthesis pathway. This, in combination with transcrip-

tome profiling using *L. japonicus* microarrays (Endo et al., 2002; Masuko et al., 2006), will accelerate identification of genes associated with the complex processes of flavonoid biosynthesis by using integrated gene-to-metabolite correlation analysis (Hirai et al., 2004; Tohge et al., 2005). Therefore, the present study will provide an experimental basis for linking flavonoid profiling and functional genomics of *L. japonicus*.

4. Experimental

4.1. Plant material

Seed coats of *L. japonicus* accessions, Miyakojima MG-20 and Gifu B-129, were scratched by glass paper and fed water for 1 d. The seedlings were transferred to a mixture of vermiculite and a commercial soil, Powersoil (mix ratio 1.3–1, Kureha Chemical Ind., Tokyo, Japan, and Kanto Hiryou Ind., Saitama, Japan), and grown for 90–120 days (from March to June) in 2005, in a greenhouse under natural sunlight. All of the leaves, stems, and flowers were harvested and pooled separately from one individual plant to serve as leaf, stem, and flower samples, respectively.

The tissues from different developmental stages were collected as follows. From 7-day-old plants, cotyledons and hypocotyls were collected separately and served as leaf and stem samples, respectively. From 14-, 21-, 30-, 60, and 90-day-old plants, leaves and stems were collected separately, and all the leaves and all the stems from one individual plant were pooled to serve as leaf and stem samples, respectively. The tissues for one experimental replicate were collected from 12 independent plants for 7-, 14-, and 21-day-old stages, or from one independent plant for 30-, 60-, and 90-day-old stages. We prepared three replicates of leaf and stem samples for each developmental stage.

4.2. Flavonoid aglycone analysis

Dry seeds, flowers, leaves and stems were immediately frozen in liquid nitrogen after harvest and stored at −80 °C. The samples were powdered with a mortar and pestle in liquid nitrogen. Flavonoids were extracted with extraction solvent (MeOH:HOAc:H_2O, 3 ml, 9:1:10 v/v) per 1 g fr. wt (Tohge et al., 2005). After removing cell debris by filtration using PVDF membrane, 0.2 μm (Whatman, Brentford, UK), the filtrate was used either for LC-FTICR/MS analysis, or evaporated and redissolved with 1.8% HCl (v/v) in *n*-BuOH, and then boiled at 95 °C for 1 h to hydrolyze it under acid conditions. After acid hydrolysis, the solvent was substituted by MeOH, and then analyzed using HPLC (Agilent 1100 series; Agilent, Palo Alto, CA, USA) with flavonol and anthocyanidin standards. Samples were applied to a reverse phase ODS column (Migtysil RP-18; 4.6 × 250 mm, 5 μm, Cica, Tokyo, Japan) eluted in 0.1% (v/v) AcOH with an increasing gradient of CH_3CN (0–45.0 min, 17–80%; 45.1–50.0 min, 95%) at a flow rate of 0.5 ml min^{-1}. The eluate was detected by a photodiode array detector in the wavelength range of 200–650 nm.

4.3. HPLC

Three HPLC conditions were tested to optimize the separation of flavonoids. In condition A, the filtered extract was applied to an ODS column, TSKgel ODS-80TM (4.6 × 150 mm, 5 μm; TOSOH, Tokyo, Japan) and eluted with a linear gradient of CH_3CN in 0.1% (v/v) TFA/H_2O. The gradient protocol was 0–40.0 min, 10–42%; 40.1–45.0 min, 95%; 45.1—52.0 min, 10% at a flow rate of 0.5 ml min^{-1}. In condition B, the filtered extract was applied to TSKgel ODS-100 V (4.6 × 250 mm, 5 μm; TOSOH), and eluted by the same protocol as in condition A. In condition C, the filtered extract was applied to TSKgel ODS-100 V (TOSOH) with a linear gradient of CH_3CN in 0.1% (v/v) HCO_2H/H_2O. The gradient protocol was 0–45.0 min, 3–42%; 45.1–50.0 min, 95%; 50.1–57.0 min, 3% at a flow rate of 0.5 ml min^{-1}.

4.4. LC-FTICR/MS

To monitor HPLC elution, a photodiode array detector was used in the wavelength range of 200–650 nm. The eluate was introduced on-line to a Finnigan LTQ-FT (Thermo Electron, Waltham, MA, USA), and positive-ion ESI-MS was performed. Nitrogen gas was used as sheath and auxiliary gas. For ionization, the capillary voltage was set at 3.5 kV, and the capillary temp. at 300 °C. A full MS scan with internal standards was performed in the range of 200–1500 (*m*/*z*) at resolution 100,000 (at *m*/*z* 400), capillary voltage 15.0 V, and tube lens offset 70.0 V. A mixture of internal standards (20 μM lidocain, 5 μM prochloraz, 2.5 μM reserpine, 3.75 μM bombesin and 1 μM aureobasidin A) dissolved in 0.1% (v/v) HCO_2H/MeOH was introduced by a post-column addition method described previously (Huhman and Sumner, 2002) at a flow rate of 5 μl min^{-1}. MS/MS fragmentation was carried out at normalized collision energy 35.0% and isolation width 4.0 (*m*/*z*).

4.5. Calibration of m/z using internal standards

Xcalibur software (Thermo Electron) was used for data acquisition and for manually browsing the acquired data. For simulating isotope peak patterns, the Qual Browser module of Xcalibur was used. For bulk-processing of FTICR/MS data, a program, MSGet, was developed using Microsoft Excel Visual Basic for Application to acquire mass spectral data in an XRAW file format from the Finnigan LTQ-FT (Thermo Electron), and transformed into a simple text file format. By selecting the scan filter and retention time range, data in each mass scan was obtained.

A Java-based program, DMP, was also developed to add information about internal standards, including their theoretical *m/z* values and observed *m/z* values, to the data obtained using MSGet. The output file was readied for analytical error correction using a computational tool, DrDMASS (http://kanaya.aist-nara.ac.jp/DrDMASS/) as described previously (Oikawa et al., 2006). The experimental *m/z* values of the internal standards were fixed to their theoretical values, and the *m/z* error calibration data were used for *m/z* compensation for all other ions in each mass scan.

4.6. Quantification of flavonoids

Flavonoid content was determined by the A_{280} and A_{520} peak areas for flavonols and anthocyanins, respectively, using calibration curves of standard compounds.

4.7. Purification of gossypetin glycoside

Dried powder of flowers from MG-20 and B-129 were extracted with extraction solvent (3 L, MeOH:HOAc:H_2O = 9:1:10). After passing through filter paper (Qualitative) (ADVANTEC, Tokyo, Japan), MEOH in the extract was removed using a rotary evaporator. The aqueous solution was extracted with hexane three times. The aqueous phase was absorbed to AMBERLITE XAD-2 resin (Organo, Tokyo, Japan) at 4 °C in the dark. Flavonoids were eluted with MeOH from the adsorbent. After evaporation, the dried eluate was dissolved in extraction solvent, and was filtered using PVDF membrane, 0.2 μm (Whatman). The filtrate was applied onto HPLC column Migtysil RP-18 (10 × 250 mm, 5 μm; Cica), and eluted with an increasing gradient of CH_3CN (0–20.0 min, 13–15%; 20.1–25.0 min, 95%; 25.1–35.0 min, 13%) in 0.1% (v/v) TFA at a flow rate of 4.0 ml min^{-1} at 30 °C. The fractions containing gossypetin glycoside were pooled, concentrated, and further purified by HPLC using the same elution protocol. Purified gossypetin glycoside was concentrated by AMBERLITE XAD-2 column and filtered using PVDF membrane 0.2 μm (Whatman).

4.8. NMR

^1H and ^{13}C NMR spectra were measured on a JEOL JNM ecp600 at 600 (^1H) and 150 MHz (^{13}C), respectively. To the best of our knowledge, full assignments of ^1H and ^{13}C NMR of gossypetin 3-*O*-β-galactoside and gossypetin 3-*O*-β-gulucoside have not been reported elsewhere.

4.8.1. Gossypetin 3-O-β-galactoside (LJ-YP1)

^1H NMR (DMSO-d_6, 600 MHz): δ 12.06 (1H, *s*, 5-OH), 10.54 (1H, *s*, 7-OH), 9.71 (1H, *s*, 4′-OH), 9.16 (1H, *s*, 3′-OH), 8.65 (1H, *s*, 8-OH), 7.74 (1H, *dd*, *J* = 8.6, 1.7 Hz, H-6′), 7.64 (1H, *d*, *J* = 1.7 Hz, H-2′), 6.82 (1H, *d*, *J* = 8.6 Hz, H-5′), 6.28 (1H, *s*, H-6), 5.36 (1H, *d*, *J* = 7.6 Hz, H-1″), 5.13 (1H, *br d*, *J* = 4.6 Hz, 2″-OH), 4.83 (1H, *br d*, *J* = 5.3 Hz, 3″-OH), 4.44 (1H, *br t*, *J* = 5.6 Hz, 6″-OH), 4.42 (1H, *br d*, *J* = 4.0 Hz, 4″-OH), 3.58 (1H, *m*, H-2″), 3.46 (1H, *m*, H-6″), 3.37 (1H, *m*, H-3″), 3.28 (1H, *m*, H-6″), H-4″ and H-5″ were buried under H_2O signal. ^{13}C NMR (DMSO-d_6, 150 MHz): δ 156.7 (C-2), 133.8 (C-3), 178.4 (C-4), 153.2 (C-5), 99.0 (C-6), 153.6 (C-7), 125.4 (C-8), 145.4* (C-9), 104.2 (C-10), 122.0 (C-1′), 116.8 (C-2′), 145.3* (C-3′), 149.0 (C-4′), 115.6 (C-5′), 122.7 (C-6′), 102.5 (C-1″), 71.8 (C-2″), 73.8 (C-3″), 68.5 (C-4″), 76.4 (C-5″), 60.7 (C-6″), * interchangeable.

4.8.2. Gossypetin 3-O-β-glucoside (LJ-YP2)

^1H NMR (DMSO-d_6, 600 MHz): δ 12.07 (1H, *s*, 5-OH), 10.52 (1H, *s*, 7-OH), 9.70 (1H, *s*, 4′-OH), 9.20 (1H, *s*, 3′-OH), 8.64 (1H, *s*, 8-OH), 6.28 (1H, *s*, H-6), 7.68 (1H, overlapped, H-2′), 7.67 (1H, overlapped, H-6′), 6.85 (1H, *d*, *J* = 8.6 Hz, H-5′), 5.54 (1H, *d*, *J* = 7.3 Hz, H-1″), 5.27 (1H, *br d*, *J* = 4.6 Hz, 2″-OH), 5.04 (1H, *br d*, *J* = 4.3 Hz, 3″-OH), 4.93 (1H, *br d*, *J* = 4.6 Hz, 4″-OH), 4.22 (1H, *br t*, *J* = 5.3 Hz, 6″-OH), 3.58 (1H, *br dd*, *J* = 11.7, 5.3 Hz, H-6″), 3.24 (2H, overlapped, H-2″ and H-3″), 3.09 (2H, overlapped, H-4″ and H-5″), H-6″ was buried under H_2O signal. ^{13}C NMR (DMSO-d_6, 150 MHz): δ 156.6 (C-2), 133.7 (C-3), 178.3 (C-4), 153.3 (C-5), 99.0 (C-6), 153.6 (C-7), 125.4 (C-8), 145.4 (C-9), 104.2 (C-10), 122.1 (C-1′), 117.0 (C-2′), 145.3 (C-3′), 149.0 (C-4′), 115.6 (C-5′), 122.4 (C-6′), 101.5 (C-1″), 74.7 (C-2″), 77.1 (C-3″), 70.5 (C-4″), 78.1 (C-5″), 61.6 (C-6″).

Acknowledgements

We thank Dr. Yoko Iijima (Kazusa DNA Research Institute) for critical reading of the manuscript. This work was performed as part of the technology development projects of the Green Biotechnology Program supported by New Energy and Industrial Technology Development (NEDO).

References

Aharoni, A., Ric de Vos, C.H., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R., Goodenowe, D.B., 2002. Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. Omics 6, 217–234.

Asamizu, E., Kato, T., Sato, S., Nakamura, Y., Kaneko, T., Tabata, S., 2003. Structural analysis of a *Lotus japonicus* genome. IV. Sequence features and mapping of seventy-three TAC clones which cover the 7.5 mb regions of the genome. DNA Res. 10, 115–122.

Asamizu, E., Nakamura, Y., Sato, S., Tabata, S., 2004. Characteristics of the *Lotus japonicus* gene repertoire deduced from large-scale expressed sequence tag (EST) analysis. Plant Mol. Biol. 54, 405–414.

Endo, M., Matsubara, H., Kokubun, T., Masuko, H., Takahata, Y., Tsuchiya, T., Fukuda, H., Demura, T., Watanabe, M., 2002. The advantages of cDNA microarray as an effective tool for identification of reproductive organ-specific genes in a model legume, *Lotus japonicus*. FEBS Lett. 514, 229–237.

Farag, M.A., Huhman, D.V., Lei, Z., Sumner, L.W., 2007. Metabolic profiling and systematic identification of flavonoids and isoflavonoids

in roots and cell suspension cultures of *Medicago truncatula* using HPLC-UV-ESI-MS and GC–MS. Phytochemistry 68, 342–354.

Graham, T.L., 1991. Flavonoid and isoflavanoid distribution in developing soybean seedling tissues and in seed and root exudates. Plant Physiol. 95, 594–603.

Hall, R.D., 2006. Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytol. 169, 453–468.

Harborne, J.B., 1969. Gossypetin and herbacetin as taxonomic markers in higher plants. Phytochemistry 8, 177–183.

Hirai, M.Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D.B., Yamazaki, Y., Kanaya, S., Nakamura, Y., Kitayama, M., Suzuki, H., Sakurai, N., Shibata, D., Tokuhisa, J., Reichelt, M., Gershenzon, J., Papenbrock, J., Saito, K., 2005. Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. J. Biol. Chem. 280, 25590–25595.

Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T., Saito, K., 2004. Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 101, 10205–10210.

Huhman, D.V., Sumner, L.W., 2002. Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry 59, 347–360.

Jay, M., Hasan, A., Voirin, B., Viricel, M.R., 1978. Les flavonoides du *Lotus corniculatus*. Phytochemistry 17, 827–829.

Kaneko, T., Asamizu, E., Kato, T., Sato, S., Nakamura, Y., Tabata, S., 2003. Structural analysis of a *Lotus japonicus* genome. III. Sequence features and mapping of sixty-two TAC clones which cover the 6.7 Mb regions of the genome. DNA Res. 10, 27–33.

Kato, T., Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Tabata, S., 2003. Structural analysis of a *Lotus japonicus* genome. V. Sequence features and mapping of sixty-four TAC clones which cover the 6.4 mb regions of the genome. DNA Res. 10, 277–285.

Kawaguchi, M., Motomura, T., Imaizumi-Anraku, H., Akao, S., Kawasaki, S., 2001. Providing the basis for genomics in *Lotus japonicus*: the accessions Miyakojima and Gifu are appropriate crossing partners for genetic analyses. Mol. Genet. Genomics 266, 157–166.

Kind, T., Fiehn, O., 2006. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. BMC Bioinformatics 7, 234.

Kouchi, H., Shimomura, K., Hata, S., Hirota, A., Wu, G.J., Kumagai, H., Tajima, S., Suganuma, N., Suzuki, A., Aoki, T., Hayashi, M., Yokoyama, T., Ohyama, T., Asamizu, E., Kuwata, C., Shibata, D., Tabata, S., 2004. Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, *Lotus japonicus*. DNA Res. 11, 263–274.

Lazar, G., Goodman, H.M., 2006. MAX1, a regulator of the flavonoid pathway, controls vegetative axillary bud outgrowth in Arabidopsis. Proc. Natl. Acad. Sci. USA 103, 472–476.

Martens, S., Mithofer, A., 2005. Flavones and flavone synthases. Phytochemistry 66, 2399–2407.

Masuko, H., Endo, M., Saito, H., Hakozaki, H., Park, J.I., Kawagishi-Kobayashi, M., Takada, Y., Okabe, T., Kamada, M., Takahashi, H., Higashitani, A., Watanabe, M., 2006. Anther-specific genes, which expressed through microsporogenesis, are temporally and spatially regulated in model legume, *Lotus japonicus*. Genes Genet. Syst. 81, 57–62.

Murch, S.J., Rupasinghe, H.P., Goodenowe, D., Saxena, P.K., 2004. A metabolomic analysis of medicinal diversity in Huang-qin (*Scutellaria baicalensis* Georgi) genotypes: discovery of novel compounds. Plant Cell Rep. 23, 419–425.

Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Sato, S., Tabata, S., 2002. Structural analysis of a *Lotus japonicus* genome. II. Sequence features and mapping of sixty-five TAC clones which cover the 6.5-mb regions of the genome. DNA Res. 9, 63–70.

Nesi, N., Debeaujon, I., Jond, C., Stewart, A.J., Jenkins, G.I., Caboche, M., Lepiniec, L., 2002. The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell 14, 2463–2479.

Oikawa, A., Nakamura, Y., Ogura, T., Kimura, A., Suzuki, H., Sakurai, N., Shinbo, Y., Shibata, D., Kanaya, S., Ohta, D., 2006. Clarification of pathway-specific inhibition by fourier transform ion cyclotron resonance/mass spectrometry-based metabolic phenotyping studies. Plant Physiol. 142, 398–413.

Pfenninger, A., Karas, M., Finke, B., Stahl, B., 2002. Structural analysis of underivatized neutral human milk oligosaccharides in the negative ion mode by nano-electrospray MS(n) (part 2: application to isomeric mixtures). J. Am. Soc. Mass Spectrom. 13, 1341–1348.

Pourcel, L., Routaboul, J.M., Kerhoas, L., Caboche, M., Lepiniec, L., Debeaujon, I., 2005. TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat. Plant Cell 17, 2966–2980.

Reynaud, J., Lussignol, M., 2005. The flavonoids of *Lotus corniculatus*. *Lotus* Newsletter 35, 75–82.

Sato, S., Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T., Tabata, S., 2001. Structural analysis of a *Lotus japonicus* genome. I. Sequence features and mapping of fifty-six TAC clones which cover the 5.4 mb regions of the genome. DNA Res. 8, 311–318.

Seshadri, T.R., Thakur, R.S., 1961. Constitution of gossytrin, a new glycoside from the flower petals of *Hibiscus sabdariffa*. J. Indian Chem. Soc. 38, 649–651.

Shimada, N., Aoki, T., Sato, S., Nakamura, Y., Tabata, S., Ayabe, S., 2003. A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso)flavonoids in *Lotus japonicus*. Plant Physiol. 131, 941–951.

Shimada, N., Sasaki, R., Sato, S., Kaneko, T., Tabata, S., Aoki, T., Ayabe, S., 2005. A comprehensive analysis of six dihydroflavonol 4-reductases encoded by a gene cluster of the *Lotus japonicus* genome. J. Exp. Bot. 56, 2573–2585.

Springob, K., Nakajima, J., Yamazaki, M., Saito, K., 2003. Recent advances in the biosynthesis and accumulation of anthocyanins. Nat. Prod. Rep. 20, 288–303.

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, M., Noji, M., Yamazaki, M., Saito, K., 2005. Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. Plant J. 42, 218–235.

Yergey, J.A., 1983. A general approach to calculating isotopic distributions for mass spectrometry. Int. J. Mass Spectrom. Ion Phys. 52, 337–349.

Zaia, J., Li, X.Q., Chan, S.Y., Costello, C.E., 2003. Tandem mass spectrometric strategies for determination of sulfation positions and uronic acid epimerization in chondroitin sulfate oligosaccharides. J. Am. Soc. Mass Spectrom. 14, 1270–1281.