Chemical Profiling of Lentil (*Lens culinaris* Medik.) Cultivars and Isolation of Compounds

APOLLINAIRE TSOPMO*,† AND ALISTER D. MUIR§

†Food Science and Nutrition Program, Department of Chemistry, Carleton University, Ottawa, Ontario K1S 5B6, Canada, and §Agriculture and Agri-Food Canada, 107 Science Place, Saskatoon, Saskatchewan S7N 0X2, Canada

A high-performance liquid chromatography method was developed to obtain fingerprints of secondary metabolites of 12 lentil cultivars grown under the same environmental condition. Extracts (100% methanol and methanol−water (1:1)) were analyzed by RP-HPLC. Full photodiode array (191−360 nm) data were collected and used for cluster analysis. Methanol and methanol−water extracts showed slightly different clustering patterns. In the dendogram of methanol extracts, CDC Richlea appeared as an isolated group, whereas Indianhead was the isolated group in methanol−water extracts. The cultivar CDC Milestone was selected for further evaluation because of the presence of three peaks (8.9, 16.7, and 32.7 min) that were absent in other cultivars or present in very small amounts. Chromatographic separations of the methanol extract afforded several compounds including the novel 4-chloro-1*H*-indole-3-*N*-methylacetamide (**13**) as well as itaconic acid (**3**), arbutin (**5**), gentisic acid 5-*O*-[β-D-apiofuranosyl-(1→2)-β-D-xylopyranoside] (**9**), and (6*S*,7*Z*,9*R*)-9-hydroxy-megastigma-4,7-dien-3-one-9-*O*-β-D-apiofuranosyl-(1→2)-β-D-glucopyranoside (**14**), which are described for the first time from lentils. Structures were determined by high-resolution NMR experiments.

KEYWORDS: Lentil; cluster analysis; kaempferol glycosides; chemical composition

INTRODUCTION

Plants and crops produce a vast number of secondary metabolites to overcome environmental stress and protect themselves from their natural enemies. Humans use some of these compounds for nutraceutical or pharmaceutical purposes. In recent years, there has been an increased interest in the comprehensive analysis of these secondary metabolites prior to lengthy and costly preparative separation procedures (*1*). Such approaches are termed "fingerprinting", as they emphasize the pattern recognition of metabolites after chromatographic profiling rather than the cataloging of specific compounds (*2*). The chemical compositions of foods sold at restaurants were analyzed using this technique (*3*). It was also used to identify novel natural compounds (*4*), to evaluate the mode of action of various herbicides in plants (*5*), and to determine metabolic changes in transgenic crops (*6*). The fingerprinting requires the development of a method (often HPLC, LC-MS, GC-MS, or NMR) that encompasses the diversity of phytochemicals of which more than 200,000 have been estimated to occur in the plant kingdom while achieving a sufficiently high analytical throughput to make the experimentation feasible (*7*). Data generated are often complex and required chemometric methods for their interpretation.

Lentil (*Lens culinaris* Medik.) is a common legume known since biblical times. Bahl et al. (*8*) said that it is probably the first of the grain legumes to be domesticated. In North America much of the acreage is in western Canada, eastern Washington, and northern Idaho, where dry growing season conditions prevail (*9*). Lentils are usually placed in two groups of subspecific rank (*10*). The first group *Lens culinaris* ssp. *macrosperma* (Baumb.) Barulina have large flattened seeds (6−9 mm diameter) and white or blue large flowers while the second group *L. culinaris* ssp. *microsperma* (Baumb.) Barulina possess small seeds (2−6 mm diameter) and flowers that are violet-blue to white or pink. The *macrosperma* types are important in the northern hemisphere, whereas the *microsperma* are principally found in southwestern and western Asia and Africa (*10*).

L. culinaris is known to contain large amounts of the oligosaccharides raffinose, stachyose, and verbascose (*11*) as well as the sugar alcohols cyclotol D-pinitol, galactopinitol, and ciceritol (*12*). It is also contains antinutritional factors such as trypsin inhibitors and hemagglutinins (*13*). Concentrations of simple phenolic components have been reported (*14*, *15*), and recently flavonol glycosides were reported from seeds (*16*).

The aim of this work was to apply the fingerprinting technique to lentil cultivars grown under the same environmental conditions in Saskatchewan, Canada, and to identify specific constituents that can be used to improve breeding or to develop novel functional foods or nutraceuticals.

EXPERIMENTAL PROCEDURES

Chemicals and Reagents. HPLC-grade water was obtained by passing RO-treated water through a Super Q water system (Millipore Canada, Mississauga, Canada) prior to use. HPLC-grade acetonitrile, methanol, hexane, and trifluoroacetic acid (TFA) were obtained from

*Corresponding author [phone (613) 520-2600, ext. 3122; fax (613) 520-3749; e-mail apollinaire_tsopmo@carleton.ca].

Table 1. Twelve Lentil Cultivars Included in the Cluster Analysis

name code	description	cultivar name	source	PGRC[a] accession no.
LGL	large green	CDC Glamis	Bailey Bro Seeds, Milden, SK	CN 105577
LGR	large green	CDC Grandora	Bailey Bro Seeds, Milden, SK	CN 105957
LSO	large green	CDC Sovereign	Bailey Bro Seeds, Milden, SK	CN 105960
LLA	large green	Laird	Simpson Seeds, Pasqua, SK	CN 3541
LVA	medium green	CDC Vantage	Bailey Bro Seeds, Milden, SK	CN 105575
LES	Persian green	Eston	Simpson Seeds, Pasqua, SK	CN 40010
LIH	black	Indian Head	Farley Seeds, Regina, SK	CN 105579
LMI	small green	CDC Milestone	Sunset Farms, Pennant, SK	CN 105576
LRC	small red	CDC Redcap	FarmPure Seeds, Moose Jaw, SK	CN 105958
LRB	small red	CDC Robin	Ardell Seeds, Vanscoy, SK	CN 105959
LRW	small red	CDC Redwing	PGRC	CN 105582
LRL	small green	CDC Richlea	PGRC	CN 105583

[a] PGRC, Plant Gene Resources of Canada.

VWR Canada. The internal standards ferulic acid and flavone (2-phenyl-4*H*-1-benzylpyran-4-one) were obtained from Sigma-Aldrich Canada and used without any further purification.

Plant Material. The seeds of the lentil cultivars were obtained from different distributors (**Table 1**) and were all grown in the same plot in summer 2007 in Saskatoon (SK, Canada). For cluster analysis experiments, 1 g of seeds from each cultivar was pulverized and extracted with hexane (25 mL) in a Swedish tube for 3 h. After vacuum filtration, each meal sample (500 mg) was extracted with 5 mL of MeOH followed by 5 mL of MeOH/H_2O (1:1) for 2 h in a water bath at 60 °C. The combined extracts were centrifuged and filtered through a 0.45 μm nylon syringe filter for RP-HPLC analysis.

HPLC Conditions. LC-DAD (diode array detection) studies were performed on an Alliance 2690 separation module equipped with vacuum solvent degassing (Waters Canada, Mississauga, Canada) and a Waters 996 photodiode array detector. Instrument control, data collection, and analysis were undertaken using Empower 2 software (Waters). Samples were analyzed on a Symmetry C_{18} column (3.0 × 150 mm, 5 μm, Waters). The mobile phases were (A) 0.05% TFA in water and (B) 0.05% TFA in acetonitrile. The injection was 10 μL for all samples and standards. The column was eluted (0.4 μL min^{-1}) with the following gradient: 0 min, 2% B; 5 min, 2% B; 25 min, 25% B, 30 min, 40% B; 40 min, 95% B; 45 min, 95% B; 55 min, 2% B; and 65 min, 2% B. Preparative HPLC was carried out on a Waters DeltaPrep 4000 system equipped with a 486 UV detector.

Initial Fractionation. Whole seeds of CDC Milestone (LMI) lentil cultivar were ground with a Wiley mill equipped with a 40 mesh screen. The flour (2.6 kg) was defatted in a Soxhlet with 8 L of hexane for 4 h, and the hexane extract was obtained by rotary evaporation (59.3 g). The defatted flour collected by Buchner filtration was extracted with 100% methanol (12 L) at 60 °C for 5 h, and the filtrate was concentrated by rotary evaporation on Buchi R220 (bath temperature < 40 °C) to yield 150.4 g of extract. The concentrated extract was diluted with water (1000 mL) and subjected to liquid–liquid extraction with ethyl acetate (3 × 600 mL) and then *n*-butanol (3 × 600 mL). Only the remaining aqueous extract was concentrated and used for preparative separation. Initial separation was performed on a DeltaPrep 4000 system, and the eluent was a linear gradient of 2–98% of solvent B (100% methanol) in solvent A (1% Acetic acid in water) at a flow rate 50 mL/min over 90 min. The UV detector was set 280 nm, and a total of 10 fractions, F1–F10, were collected.

Purification of Constituents. All Sephadex LH-20 gel permeation separations were performed on a 4.9 × 60 cm (i.d. × L) open column. Fractions F1 and F2 from DeltaPrep separation were combined and separated by gel permeation on Sephadex LH-20 that had been soaked for 3 h in a mixture of methanol/water (1:1). Fractions from the column were monitored by TLC on precoated silica gel 60 (CHCl₃/MeOH/H₂O 9:7:1) and the plates visualized under a UV lamp at 280 nm or by spraying with a solution of 10% sulfuric acid followed by heating with a heat gun. NMR analysis showed that these fractions contained stachyose (**1**), 4-*O*-methyl-pinitol-2-α-D-galactoside (**2**) as well as raffinose, sucrose, and D-glucose. Fraction F3 was fractionated on Sephadex LH-20 (H₂O/MeOH 4:6) and then by open column chromatography (2.5 × 30 cm) on silica gel with CHCl₃/MeOH (8:2) to afford itaconic acid or methylenebutanedioic acid (**3**). Gel permeation (MeOH/H₂O 6:4) separation of fraction 4 afforded a

mixture of two glycoside compounds as observed on NMR spectra. The mixture (60 mg) was then treated with 10 mL of a pyridine/acetic anhydride (1:1) solution and stirred at room temperature for 2 h. The reaction mixture was dried by rotary evaporation and loaded onto a silica gel column (1.5 × 50 cm). Elution with CHCl₃/MeOH (98:2) yield two compounds identified as acetylated derivatives of uridine (**4**) and arbutin (**5**). Fraction 5 was gel permeated on Sephadex LH-20 in 100% methanol, and the following compounds were obtained: trigonelline (**6**), *epi*-dihydrophaseic acid 4′-*O*-β-D-glucopyranoside (**7**), hypaphorine (**8**), and gentisic acid 5-*O*-[β-D-apiofuranosyl-(1→2)-β-D-xylopyranoside] (**9**). Repeated gel permeation separation of fraction F6 with MeOH/H₂O (8:2) as eluent yielded kaempferol 3-*O*-β-D-glucopyranosyl(1→2)-*O*-[α-L-rhamnopyranosyl(1→6)]-β-D-glucopyranoside-7-*O*-α-L-rhamnopyranoside (**10**), kaempferol 3-*O*-β-D-glucopyranosyl(1→2)-*O*-[α-L-rhamnopyranosyl(1→6)]-β-D-galactopyranoside-7-*O*-α-L-rhamnopyranoside (**11**), and kaempferol 3-*O*-β-D-glucopyranosyl(1→2)-*O*-β-D-galactopyranoside-7-*O*-α-L-rhamnopyranoside (**12**). Fraction 7 was subjected to Sephadex LH-20 gel permeation with MeOH/H₂O (9:1) to afford a subfraction. This was further purified on a silica gel column (2.5 × 30 cm) with mixtures of CHCl₃/MeOH/H₂O 10:5:0.5 and 10:7:1 to yield 4-chloro-1*H*-indole-3-*N*-methyl-lacetamide (**13**), (6*S*,7*Z*,9*R*)-9-hydroxymegastigma-4,7-dien-3-one-9-*O*-β-D-apiofuranosyl-(1→2)-β-D-glucopyranoside (**14**) and astragalin-7-*O*-α-L-rhamnopyranoside (**15**). Fractions 8–10 were combined and purified on a silica gel column (2.5 × 30 cm) eluted with CHCl₃/MeOH mixtures 9:1 and 8:2, and the monoglycoside 7-*O*-α-L-rhamnopyranosyloxy-3,4′,5-trihydroxyflavone (**16**) was obtained.

Mass Spectrometry. Positive ion electrospray ionization (ESI) mass spectra were obtained with a benchtop tandem quadrupole mass spectrometer (Quattro *micro* API 2000, Micromass Ltd., U.K.) equipped with an atmospheric pressure ESI source interfaced directly connected to a Waters Alliance 2690 separations module. MS parameters were as follows: capillary voltage, 3.2 kV; source temperature, 80 °C; desolvation temperature, 200 °C; cone gas (nitrogen), 50 L/h; desolvation gas (nitrogen), 500 L/h; and electron multiplier, 650 V. The instrument was controlled by Micromass MassLynx software version 4.0. A C_{18} Symmetry column (2.1 mm × 150 mm, 5 μm) was used for all analyses. The mobile phase (0.2 mL/min) consisted of 0.1% formic acid in water (solvent A) and 100% acetonitrile (solvent B). The gradient program was 98% solvent A and 2% solvent B up to 2% A and 98% B in 65 min. Collision-induced dissociation (CID) experiments were conducted in the positive ion mode with argon and collision energies of 30–35 eV. Quasimolecular ions (MH⁺) were used to generate the daughters' ion spectra.

NMR Spectroscopy. Spectra were obtained at room temperature for all samples in methanol-d_4, dimethyl sulfoxide-d_6, or a mixture of these containing about 5% deuterated water (D₂O) with a Bruker AVANCE 500 spectrometer (TopSpin 1.3 software) equipped with a Bruker BioSpin 5 mm CryoProbe with z-gradient. Chemical shifts (δ in ppm) were referenced to solvent resonances at 3.31 and 49.0 ppm for ¹H and ¹³C, respectively, for samples run in methanol-d_4 or 2.50 (¹H) and 39.5 (¹³C) for samples in DMSO-d_6.

Cluster Analysis of Samples. The HPLC data (retentions time, areas) were exported into Microsoft Excel program and the retention times aligned. Each peak was assigned a score of 1, and any missing peak
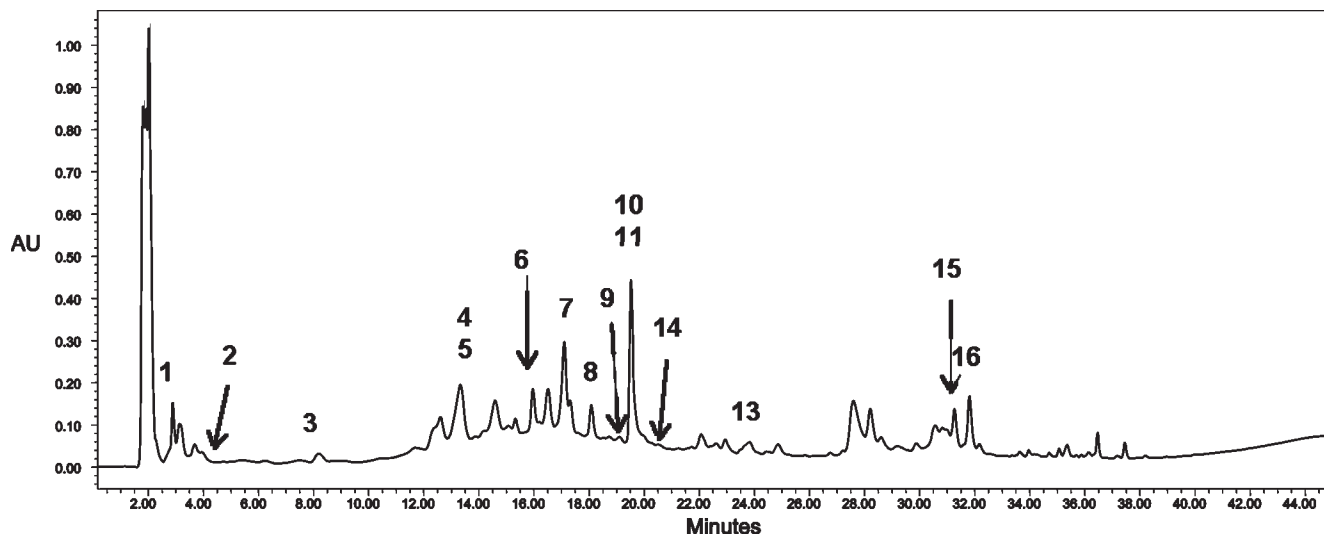
Article

J. Agric. Food Chem., Vol. 58, No. 15, 2010   **8717**

Figure 1. HPLC-DAD chromatogram of 100% LMI (CDC Milestone) lentil cultivar. Numbers correspond to compounds identified after separation and NMR analysis.

(present in one chromatogram but missing in another) was assigned a score of 0. The Excel data table was imported into NTSYSpc Software version 2.1 (Numerical Taxonomy SYStem for pc) and the SIMQUAL (SIMilarity for QUALitative) module applied. The final clustering was performed using the SAHN (sequential agglomerative hierarchical nested) module (*17*). The results are represented graphically in **Figures 2** and **3**.

RESULTS AND DISCUSSION

Chemical Profiling Using HPLC Analysis. The usefulness of the cluster analysis depends upon the reproducibility and consistency of the data that are used. We evaluated a number of compounds as suitable internal standards to be used to normalize the chromatographic data, and ferulic acid was chosen. The retention time of ferulic acid in all chromatograms was calibrated to 24.4 min. The HPLC gradient we used gave a relatively good separation of peaks present in extracts of the cultivars. We recorded full photodiode array spectral profiles and analyzed for the best wavelength for routine data collection. Many of the compounds in extracts did not have significant absorption above 220 nm; therefore, it was determined that a Maxplot would have to be used in analysis. In the MaxPlot the chromatogram is a composite of the maximum absorption at any given point in time within the specified wavelength range, which in this case was 191−360 nm.

Several extractions were made to determine the best method of screening for most of the compounds present in cultivars. These included 40, 50, and 60 °C, as well as room temperature, and five solvents, 100 and 50% of methanol and ethanol and then 100% water. All samples were run and UV chromatograms (191−360 nm) compared. Extracts that displayed most of the peaks were obtained with MeOH and MeOH/H_2O (1:1) at 60 °C.

Twelve commercial cultivars (**Table 1**) were subjected to extractions, and the 100 methanol and 50% methanol extracts were analyzed by reverse phase chromatography. The HPLC data for all 12 samples were then imported into Microsoft Excel 2003, and each peak was assigned a score of 1. The retention times were then aligned, and any missing peak (i.e., a peak present in one chromatogram but missing in another) was assigned a score of 0. The Excel data table was then imported into NTSYSpc 2.1, and cluster analysis was performed using the similarity (qualitative) matrix SIMQUAL. The output is in the form of a tree-matrix called a dendogram. This tree matrix is an efficient method for describing a system of nested clusters in which each one has a corresponding level (*3*). A sample UV chromatogram is shown in **Figure 1**.

The dendogram of the 100% MeOH extracts is shown in **Figure 2** (see also the Supporting Information). Group A appeared isolated and composed of only one green lentil cultivar, CDC Richlea. Group B can be broken down into subgroups, D and C. The only cultivar present in subgroup C was CDC Milestone, another green lentil. Subgroup D can be further broken down into subgroup E, composed of two red (CDC Redwing, CDC Robin) and one green (CDC Vantage) lentil, and subgroup F, comprising five green (CDC Glamis, CDC Sovereign, Laird, CDC Grandora, Eston), one red (CDC Redcap), and one black (Indianhead) lentil. Although the red, green, and black lentils have different pigments, the clustering obtained appeared to reflect more than the simple seed coat color as the cultivars did not cluster according to seed coat color. Among the green lentils there was also a clear difference between the two small green (CDC Richlea, CDC Milestone) and the remaining five green lentils.

Analysis of the dendogram of the MeOH/H_2O extracts (**Figure 3**) showed as in the case of MeOH extracts an isolated group A composed this time of the black lentil, Indianhead. Group B may further be divided into different subgroups as shown in **Figure 3**. Again, the clustering was not related solely to the seed coat.

Identification of Compounds from CDC Milestone. Cluster analysis and evaluation of LC chromatograms showed that the cultivar CDC Milestone (LMI) had three peaks in its HPLC chromatogram at retention times of 8.9, 16.7, and 32.7 min that were absent in other cultivars or present in very small amounts. We then selected LMI for large-scale extraction and identification of its chemical constituents. Sixteen compounds (**Figure 4**) were obtained and characterized from 100% methanol extract that had been treated as described under Experimental Procedures on the basis of interpretation of NMR and MS spectra.

The most abundant fractions were eluted early from the reverse phase preparative column, and following gel permeation and NMR analysis stachyose (**1**) and pinitol 2-*O*-α-D-galactoside (**2**), raffinose, sucrose, and D-glucose were identified. There are well-characterized constituents of lentils and other dried legumes (*18−20*). Compound **3** was obtained as an amorphous powder. It was identified on the basis ^{1}H and ^{13}C NMR data as itaconic acid (**3**). This compound is often obtained by fermentation of carbohydrates and is used in the preparation of acrylic fibers and rubbers, to reinforce glass fiber, and in artificial diamonds and lenses (*21,22*). Compounds **4** and **5** were identified as uridine and arbutin,
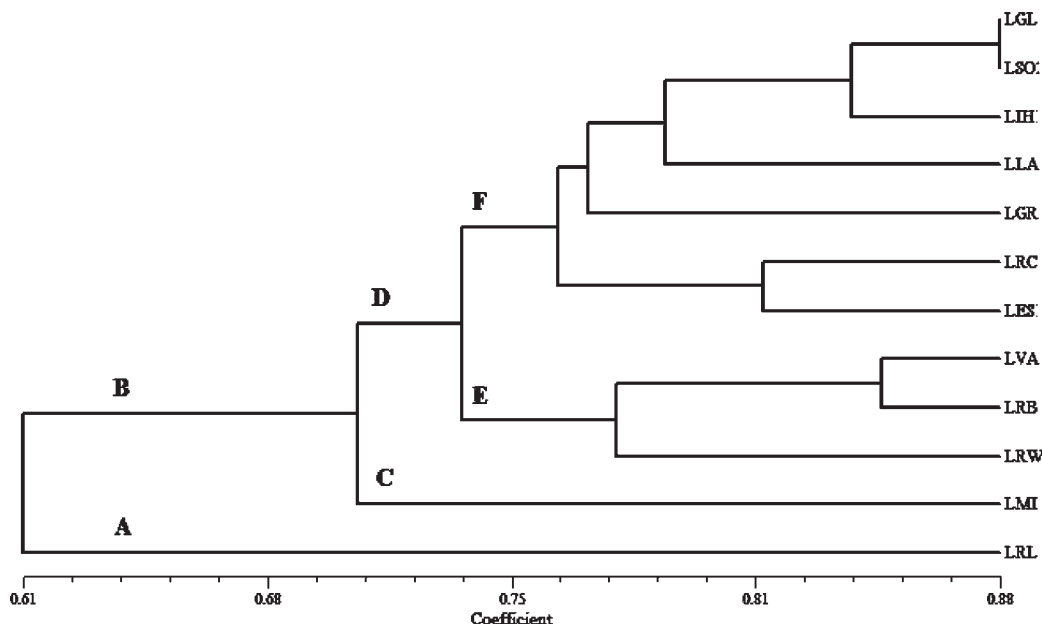
Figure 2. Dendogram of the methanol extracts of the12 commercial lentil cultivars. LGL, Glamis; LGR, Grandora; LSO, Sovereign; LLA, Laird; LVA, Vantage; LES, Eston; LIH, Indian Head; LMI, Milestone; LRC, Redcap; LRB, Robin; LRW, Redwing; LRL, Richlea.

Figure 3. Dendogram of the methanol/water extracts of the12 commercial lentil cultivars. LGL, Glamis; LGR, Grandora; LSO, Sovereign; LLA, Laird; LVA, Vantage; LES, Eston; LIH, Indian Head; LMI, Milestone; LRC, Redcap; LRB, Robin; LRW, Redwing; LRL, Richlea.
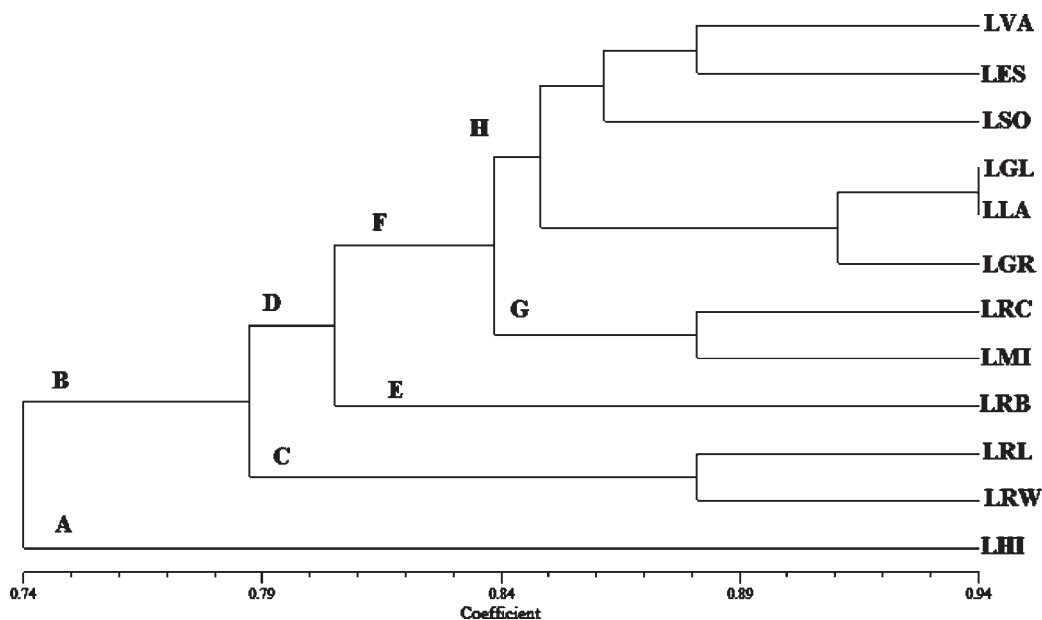
respectively, on the basis of interpretation of MS and NMR spectra of their acetylated derivatives. Uridine (**4**) is a nucleoside formed by reaction between uracil and ribose, and it is widely distributed in nature (*23*). Arbutin (**5**) has been isolated from many plants but not from lentils. Its hydrolysis and oxidation of the resultant 1,4-benzenediol are responsible for leaf blackening in many dead plants (*24*, *25*).

Several flavonol glycosides were isolated and characterized. Compounds **10** ($[M + H]^+$, m/z 903]), **11** ($[M + H]^+$, m/z 903]), **12** ($[M + H]^+$, m/z 759]), **15** ($[M + H]^+$, m/z 597]), and **16** ($[M + H]^+$, m/z 435]) were *O*-glycosides of kaempferol according to analysis of the daughters' scan (MS/MS) spectra that showed a peak at m/z 287 corresponding to the ($[M + H]^+$ of kaempferol as sequentially hexose (M − 162) and deoxyhexose (M − 146) were lost. NMR data showed that two of them were the tetraglycosides

kaempferol 3-*O*-β-D-glucopyranosyl(1→2)-*O*-[α-L-rhamnopyranosyl(1→6)]-β-D-glucopyranoside-7-*O*-α-L-rhamnopyranoside (**10**) and kaempferol 3-*O*-β-D-glucopyranosyl(1→2)-*O*-[α-L-rhamnopyranosyl(1→6)]-β-D-galactopyranoside-7-*O*-α-L-rhamnopyranoside (**11**). These tetraglycosides were recently obtained from lentils by Taylor et al. (*16*) as an unresolved mixture. In this study we successfully separated them by repeated gel permeation on Sephadex LH-20 in 100% MeOH. The other glycosides were identified as kaempferol 3-*O*-β-D-glucopyranosyl(1→2)-*O*-β-D-galactopyranoside-7-*O*-α-L-rhamnopyranoside (**12**) (*26*), kaempferol 3-*O*-β-D-glucopyranosyl-7-*O*-α-L-rhamnopyranoside (**15**) (*27*), and kaempferol 7-*O*-α-L-rhamnopyranoside (**16**) (*28*). They have been found in lentils and other plants.

Two compounds containing quaternary nitrogen were isolated. The first was identified as trigonelline (**6**), a plant hormone
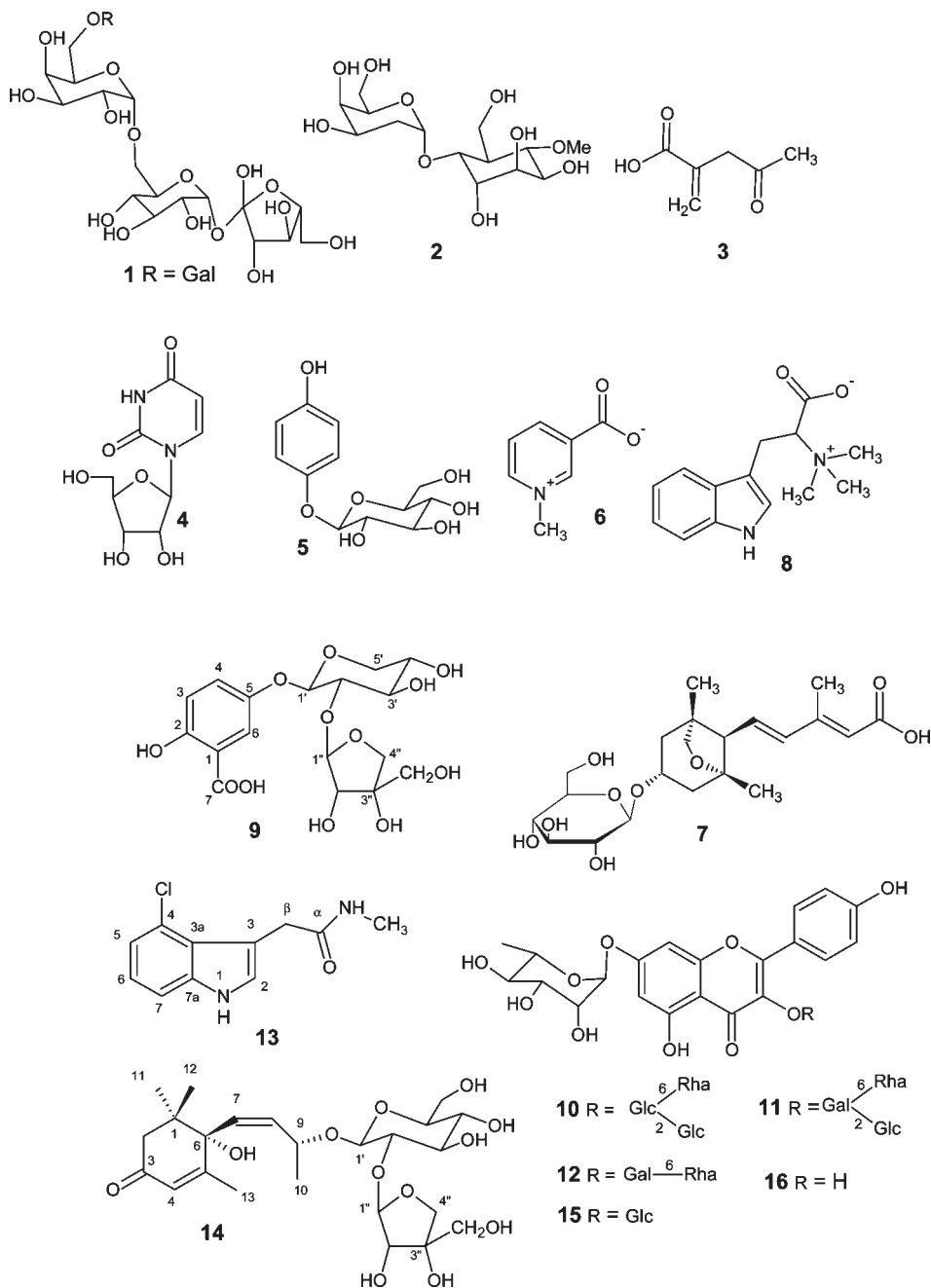
Article

J. Agric. Food Chem., Vol. 58, No. 15, 2010   **8719**

Figure 4. Chemical structures of constituents obtained from CDC Milestone cultivar.

that had been shown to possess antihyperglycaemic, antibacterial, and antifungal properties (*29, 30*); the second was identified as hypaphorine (**8**), a feeding deterrent for a seed-eating rodent previously identified in the seeds of lentils and other species (*31, 32*). Compound **7** was found to be *epi*-dihydrophaseic acid β-D-glucoside, a chemical derived from the plant hormone abscinic acid (*33, 34*).

Compound **9** was obtained as a colorless powder. Its MS spectra showed pseudomoleuclar ions at *m/z* 440 [M + Na + H]$^+$ and 417 [M − H]$^-$. The molecular formula $C_{17}H_{22}O_{12}$ was deduced from analysis of ^{13}C and HMQC spectra. Further interpretation of ^{13}C NMR and HMQC spectra of **9** showed the presence of one COOH, three CH_2, nine CH, and four quaternary C atoms. ^1H and ^{13}C NMR data (**Table 2**) of the aglycones were similar to those reported for gentisic acid 2-*O*-β-D-glucopyranose (*35*) and other gentisic glycosides (*36, 37*). The NMR spectra showed that **9** contained two sugar moieties with anomeric signals

appearing at δ_H 5.45 (*d, J* = 1.2 Hz), 4.79 (*d, J* = 7.2 Hz), and δ_C 110.8 and 101.2. HMBC correlations showed that the sugars were xylose and apiose. The xylose residue was attached to the aglycone at C-5, as confirmed by the HMBC cross-correlation peak between H-1′ (δ 4.79) and carbon C-5 (δ 150.7). Other cross-correlation peaks showed that apiose was attached to C-2′ of the xylose moiety. Therefore, compound **9** was identified as 5-*O*-[β-apiosyl-(1→2)-*O*-β-xylopyranosyl]gentisic acid. This compound is described for the first time in lentil, although it has just been found recently in *Sapatholbus suberectus* (*38*).

Compound **14** was isolated as a colorless powder. ^1H and ^{13}C NMR (**Table 2**) showed the presence of two sugar moieties. One was identified as apiose (δ_{H-1} 5.36, d *J* = 1.4 Hz, δ_{C-1} 110.7) by comparison with data of compound **9**, whereas the other was identified as glucose on the basis of characteristic chemical shifts. The aglycone was found to be a megastigmane derivative with double bonds at C-4 and C-7 positions. ^1H NMR data showed

Table 2. ¹H (500 MHz) and ¹³C (125 MHz) NMR Data (δ, Multiplicity, *J*) for Compounds **9** and **14** in CD₃OD with Solvent Signals (δ_H 3.31 and δ_C 49.0) as Reference

no.	9 C	9 H	14 C	14 H
1	119.1		42.4	
2	150.7		50.7	2.50 (1H, d, *J* = 17.0)
				2.12 (1H, d, *J* = 17.0)
3	117.6	6.71 (1H, d, *J* = 8.8)	201.2	
4	123.7	7.00 (1H, dd, *J* = 2.7, 8.8)	127.1	5.85 (1H, s)
5	158.1		167.3	
6	119.1	7.55 (1H, d, *J* = 2.7)	80.7	
7	175.4		131.5	5.83 (1H, d, *J* = 4.2)ᵃ
8			135.2	5.82 (1H, d, *J* = 4.2)
9			76.8	4.39 (1H, m)
10			21.1	1.27 (3H, d, *J* = 7.0)
11			23.4	1.03 (3H, s)
12			24.7	1.03 (3H, s)
13			19.5	1.91 (3H, s)
1′	101.2	4.79 (1H, d, *J* = 7.2)	101.2	4.37 (1H, d, *J* = 7.7)
2′	78.5	3.57 (1H, m)	79.2	3.31 (1H, m)
3′	78.4	3.41 (1H, m)	78.6	3.43 (1H, t, *J* = 8.9)
4′	71.2	3.59 (1H, m)	71.7	3.24 (1H, t, *J* = 8.7)
5′	66.0	3.58 (2H, m)	77.9	3.21 (1H, m)
6′			62.8	3.83 (1H, dd, *J* = 1.7, 11.5)
				3.63 (1H, dd, *J* = 5.2, 11.5)
1″	110.8	5.45 (1H, d, *J* = 1.2)	110.7	5.36 (1H, d, *J* = 1.4)
2″	78.2	3.96 (1H, d, *J* = 1.2)	78.0	3.91 (1H, d, *J* = 1.2)
3″	80.8		80.7	
4″	75.3	4.05 (1H, d, *J* = 9.6)	75.3	4.01 (1H, d, *J* = 10.5)
		3.78 (1H, d, *J* = 9.6)		3.72 (1H, d, *J* = 10.5)
5″	66.8	3.90 (1H, d, *J* = 4.5)	66.1	3.61 (2H, d, *J* = 2.2)
		3.88 (1H, d, *J* = 4.5)		

ᵃ Second coupling constant not obtained due to overlapping.

Table 3. ¹H (500 MHz) and ¹³C (125 MHz) NMR Data (δ, Multiplicity, *J*) for Compound **13** in CD₃OD with Solvent Signals (δ_H 3.31 and δ_C 49.0) as Reference

no.	C	H
2	128.0	7.22 (1H, s)
3	108.2	
3a	124.8	
4	126.0	
5	120.8	7.01 (1H, d, *J* = 7.2)
6	123.1	7.03 (1H, t, *J* = 7.3)
7	111.8	7.29 (1H, dd, *J* = 7.2, 1.5)
7a	139.7	
α	171.1	
β	25.3	3.82 (1H, d, *J* = 11.7)
		3.37 (1H, d, *J* = 11.7)
CH₃	52.9	3.30 (3H, s)

that the C-7 had a *cis* conformation: H-7 (δ 5.83), H-8 (δ 5.82), and $J_{\text{H-7/H-8}}$ = 4.2 Hz. Analysis of other spectra showed that **14** was (6*S*,7*Z*,9*R*)-9-hydroxymegastigma-4,7-dien-3-one-9-*O*-β-D-apiofuranosyl-(1→2)-β-D-glucopyranoside, a substance reported in the literature (*39*) but isolated here for the first time from lentils.

The ESI-MS spectrum of compound **13** showed two protonated ions peaks at *m/z* 255 and 227 with a ratio of 3:1 indicating the presence of a chloride group. The ¹³C NMR spectrum (**Table 3**) displayed the presence of 11 carbon atoms, and the HMQC direct correlation spectra showed that there were one carbonyl, one methyl, one methylene, four sp² methine groups, and four quaternary carbons. In the HMBC spectrum, important correlations were observed from the methylene groups δ_H 3.82/3.37 (d, *J* = 11.7 Hz) and the carbonyl at δ_C 171.1 and C-2

(δ 128.0), C-3 (δ 108.2), and C-3a (δ 124.8) as well as from H-7 (δ 7.29) to C-3a, C-5, C-6, and C-7a. This compound was therefore characterized as 4-chloro-1*H*-indole-3-*N*-methylacetamide (**13**), which is a novel compound.

Cluster analysis was used to classify the lentil cultivars on the basis of peaks detected on HPLC chromatograms. One cultivar that appeared to have some distinct peaks was selected for large-scale extraction and separation of its constituents. Using mass spectrometry and NMR, 16 compounds were identified. One is novel, and three are isolated for the first time in lentils. Two of the compounds specific to the CDC Milestone cultivar were identified as itaconic acid (**3**) at a retention time (RT) of 8.9 min and kaempferol monoglycoside (**16**) at RT 32.7 min. Cluster analysis may effectively be used for metabolic profiling of crop or food components before separation is attempted.

Supporting Information Available: Additional tables. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Von Roepenack-Lahaye, E.; Degenkolb, T.; Zerjeski, M.; Franz, M.; Roth, U.; Wessjohann, L.; Schmidt, J.; Scheel, D.; Clemens, S. Profiling of *Arabidopsis* secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. *Plant Physiol.* **2004**, *134*, 548–559.

(2) Wagner, C.; Sefkow, M.; Kopka, J. Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. *Phytochemistry* **2003**, *62*, 887–900.

(3) Da Silva Torres, E. A. F.; Garbelotti, M. L.; Moita Neto, J. M. The application of hierarchical clusters analysis to the study of the composition of foods. *Food Chem.* **2006**, *99*, 622–629.

(4) Wolfender, J.-L.; Rodriguez, S.; Hostettmann, K. Liquid chromatography coupled to mass spectrometry and nuclear magnetic resonance spectroscopy for the screening of plant constituents. *J. Chromatogr., A* **1998**, *794*, 299–316.

(5) Aranibar, N.; Singh, B. K.; Stockton, G. W.; Ott, K.-H. Automated mode of action detection by metabolic profiling. *Biochem. Biophus. Res. Commun.* **2001**, *286*, 150–155.

(6) Noteborn, H. P. J. M.; Lommen, A.; van der Jagt, R. C.; Weseman, J. M. Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *J. Biotechnol.* **2000**, *77*, 103–114.

(7) Fiehn, O. Metabolomics − the link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**, *48*, 155–171.

(8) Bahl, P. N.; Lal, S.; Sharma, B. M. An overview of the production and problems in southeast Asia. In *Lentil in South Asia*, Proceedings of the Seminar on Lentils in South Asia; Erskine, W., Saxena, M. C., Eds.; ICARDA: Aleppo, Syria, 1993; pp 1−10.

(9) Duke, J. A. In *Handbook of Legumes of World Economic Importance*; Plenum Press: New York, 1981; p 345.

(10) Muehlbauer, F. J.; Cubero, J. I.; Summerfield, R. J. Lentil (*Lens culinaris* Medic.). In *Grain Legume Crops*; Summerfield, R. J., Roberts, E. H., Eds.; Collins: London, U.K., 1985; pp 266−311.

(11) Gulewicz, P.; Ciesiolka, D.; Frias, J.; Vidal-Valverde, C.; Frejnagel, S.; Trojanowska, K.; Gulewicz, K. Simple method of isolation and purification of α-galactosides from legumes. *J. Agric. Food Chem.* **2000**, *48*, 3120–3123.

(12) Quemener, B.; Brillouet, J.-M. Ciceritol, a pinitol digalactoside form seeds of chickpea, lentil and white lupin. *Phytochemistry* **1983**, *22*, 1745–1751.

(13) Frias, J.; Diaz-Pollan, C.; Hedley, C. L.; Vidal-Valverde, C. Evolution of trypsin inhibitor activity during germination of lentils. *J. Agric. Food Chem.* **1995**, *43*, 2231–2234.

(14) Sosulski, F. W.; Dabrowski, K. J. Composition of free and hydrolyzable phenolic acids in the flours and hulls of ten legume species. *J. Agric. Food Chem.* **1984**, *32*, 131–133.

(15) Bartolome, B.; Estrella, I.; Hernandez, T. Changes in phenolic compounds in lentils (*Lens culinaris*) during germination and fermentation. *Z. Lebensm. Unters. Forsch. A* **1997**, *205*, 290–294.

(16) Taylor, W. G.; Fields, P. G.; Sutherland, D. H. Fractionation of lentil seeds (*Lens culinaris* Medik.) for insecticidal and flavonol tetraglycoside components. *J. Agric. Food Chem.* **2007**, *55*, 5491–5498.

(17) Sneath, P. H. A.; Sokal, R. R. In *Numerical Taxonomy: The Principles and Practice of Numerical Taxonomy*; Freeman: San Francisco, CA, 1973; p 573.

(18) Wang, N. Effect of variety and crude protein content on dehulling quality and on the resulting chemical composition of red lentil (*Lens culinaris*). *J. Agric. Food Chem.* **2008**, *56*, 885–890.

(19) Abdelgawad, A. S. Effect of domestic processing on oligosaccharide content of some dry legume seeds. *Food Chem.* **1993**, *46*, 25–31.

(20) Quemener, B.; Brillouet, J. M. Ciceritol, a pinitol digalactoside from seeds of chickpea, lentil and white lupin. *Phytochemistry* **1983**, *22*, 1745–1751.

(21) Reddy, C. S. K.; Singh, R. P. Enhanced production of itaconic acid from corn starch and market refuse fruits by genetically manipulated *Aspergillus terreus* SKR10. *Bioresour. Technol.* **2002**, *85*, 69–71.

(22) Yahiro, K.; Shibata, S.; Jia, S.-R.; Park, Y.; Okabe, M. Efficient itaconic acid production from raw corn starch. *J. Ferment. Bioeng.* **1997**, *84*, 375–377.

(23) Leinhos, V.; Krauss, G.-J.; Glund, K. Evidence that a part of cellular uridine of a tomato (*Lycopersicon esculentum*) cell suspension culture is located in the vacuoles. *Plant Sci.* **1986**, *47*, 15–20.

(24) Suau, R.; Cuevas, A.; Valpuesta, V.; Reid, M. S. Arbutin and sucrose in the leaves of the resurrection plant *Myrothamnus flabellifolia*. *Phytochemistry* **1991**, *30*, 2555–2556.

(25) Manju, M.; Varma, R. S.; Parthasarathy, M. R. New arbutin derivatives from leaves of *Grevillea robusta* and *Hakea saligna*. *Phytochemistry* **1977**, *16*, 793–794.

(26) Spanou, C.; Bourou, G.; Dervishi, A.; Aligiannis, N.; Angelis, A.; Komiotis, D.; Skaltsounis, A. L.; Kouretas, D. Antioxidant and chemopreventive properties of polyphenolic compounds derived from Greek legume plant extracts. *J. Agric. Food Chem.* **2008**, *56*, 6967–6976.

(27) Berrahal, D.; Kabouche, A.; Kabouche, Z.; Bruneau, C. Flavonoid glycosides from *Reseda villosa* (Resedaceae). *Biochem. Syst. Ecol.* **2006**, *34*, 777–779.

(28) Qin, L.; Ming, L.; Mabry, T. J.; Dixon, R. A. Flavonol glycosides from *Cephalocereus senilis*. *Phytochemistry* **1994**, *36*, 229–231.

(29) Rozan, P.; Kuo, Y.-H.; Lambein, F. Amino acids in seeds and seedlings of the genus *Lens*. *Phytochemistry* **2001**, *58*, 281–289.

(30) Chopra, S.; Ahmad, F. J.; Khar, R. K.; Motwani, S. K.; Mahdi, S.; Iqbal, Z.; Talegaonkar, S. Validated high-performance thin-layer chromatography method for determination of trigonelline in herbal extract and pharmaceutical dosage form. *Anal. Chim. Acta* **2006**, *577*, 46–51.

(31) Hofinger, M.; Monseur, X.; Pais, M.; Jarreau, F. X. Further confirmation of the presence of indolylacrylic acid in lentil seedlings and identification of hypaphorine as its precursor. *Phytochemistry* **1975**, *14*, 475–477.

(32) Janzen, D. H.; Lynn, D. G.; Fellows, L. E.; Hallwachs, W. The indole alkaloid, hypaphorine and *Pterocarpus* seed protection. *Phytochemistry* **1982**, *21*, 1035–1037.

(33) Del Refugio Ramos, M.; Jerz, G.; Villanueva, S.; López-Dellamary, F.; Waibel, R.; Winterhalter, P. Two glucosylated abscisic acid derivates from avocado seeds (*Persea americana* Mill. Lauraceae cv. Hass). *Phytochemistry* **2004**, *65*, 955–962.

(34) Zheng, Z. P.; Chen, S. B.; Wang, S. Y.; Wang, X. C.; Cheng, K. W.; Wu, J. J.; Yang, D. J.; Wang, M. F. Chemical components and tyrosinase inhibitors from the twigs of *Artocarpus heterophyllus*. *J. Agric. Food Chem.* **2009**, *57*, 6649–6655.

(35) El-Mousallamy, A. M. D.; Hussein, S. A. M.; Merfort, I.; Nawwar, M. A. M. Unusual phenolic glycosides from *Cotoneaster orbicularis*. *Phytochemistry* **2000**, *53*, 699–704.

(36) Stochmal, A.; Kowalska, I.; Janda, B.; Perrone, A.; Piacente, S.; Oleszek, W. Gentisic acid conjugates of *Medicago truncatula* roots. *Phytochemistry* **2009**, *70*, 1272–1276.

(37) Fayos, J.; Bellés, J. M.; López-Gresa, M. P.; Primo, J.; Conejero, V. Induction of gentisic acid 5-*O*-[β]-D-xylopyranoside in tomato and cucumber plants infected by different pathogens. *Phytochemistry* **2006**, *67*, 142–148.

(38) Zhang, S. W.; Xuan, L. J. New phenolic constituents from the stems of *Spatholobus suberectus*. *Helv. Chim. Acta* **2006**, *89*, 1241–1245.

(39) Wu, J.-F.; Chen, S.-B.; Gao, J.-C.; Wu, L.-J.; Chen, S.-L.; Tu, P.-F. Megastigmane glycosides from *Polygala hongkongensis* hemsl. *Chem. Res. Chinese Univ.* **2007**, *23*, 530–532.