Classification of Ilex species based on metabolomic fingerprinting using nuclear magnetic resonance and multivariate data analysis

Choi, Y.H.; Sertic, S.; Kim, H.K.; Wilson, E.G.; Michopoulosa, F.; Lefeber, A.W.M.; ... ; Verpoorte, R.

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Classification of *Ilex* Species Based on Metabolomic Fingerprinting Using Nuclear Magnetic Resonance and Multivariate Data Analysis

YOUNG HAE CHOI,† SARAH SERTIC,† HYE KYONG KIM,† ERICA G. WILSON,† FILIPPOS MICHOPOULOS,† ALFONS W. M. LEFEBER,§ CORNELIS ERKELENS,§ SERGIO D. PRAT KRICUN,# AND ROBERT VERPOORTE*,†

Division of Pharmacognosy, Section Metabolomics, Institute of Biology, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands; Division of NMR, Institute of Chemistry, Gorlaeus Laboratories, P.O. Box 9502, 2300 RA Leiden, The Netherlands; and EEA Cerro Azul INTA, Misiones, Argentina

The metabolomic analysis of 11 *Ilex* species, *I. argentina*, *I. brasiliensis*, *I. brevicuspis*, *I. dumosa* var. *dumosa*, *I. dumosa* var. *guaranina*, *I. integerrima*, *I. microdonta*, *I. paraguariensis* var. *paraguariensis*, *I. pseudobuxus*, *I. taubertiana*, and *I. theezans*, was carried out by NMR spectroscopy and multivariate data analysis. The analysis using principal component analysis and classification of the ^{1}H NMR spectra showed a clear discrimination of those samples based on the metabolites present in the organic and aqueous fractions. The major metabolites that contribute to the discrimination are arbutin, caffeine, phenylpropanoids, and theobromine. Among those metabolites, arbutin, which has not been reported yet as a constituent of *Ilex* species, was found to be a biomarker for *I. argentina*, *I. brasiliensis*, *I. brevicuspis*, *I. integerrima*, *I. microdonta*, *I. pseudobuxus*, *I. taubertiana*, and *I. theezans*. This reliable method based on the determination of a large number of metabolites makes the chemotaxonomical analysis of *Ilex* species possible.

KEYWORDS: Metabolomic analysis; *Ilex* species; arbutin; phenylpropanoids; NMR; principal component analysis; classification

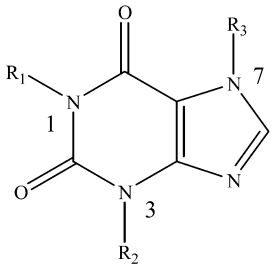
INTRODUCTION

The genus *Ilex* (Aquifoliaceae) comprises more than 500 species of dioecious trees and shrubs distributed throughout temperate and tropical regions of the world. The main areas of extant diversification are East Asia and South America. Among the species, yerba mate, a tea-like infusion of *Ilex paraguariensis* St.-Hill., is drunk in many parts of South America (*1*, *2*). It is also consumed socially in the Middle East by the Druze of Lebanon, Syria, and the Golan Heights in northern Israel. An average of 300 000 t of mate is produced each year in South America. In addition to its standing as a popular beverage, mate is used as a tonic, diuretic, and a stimulant to reduce fatigue and aid gastric function in herbal medicine systems throughout South America and other parts of the world (*3*−*5*).

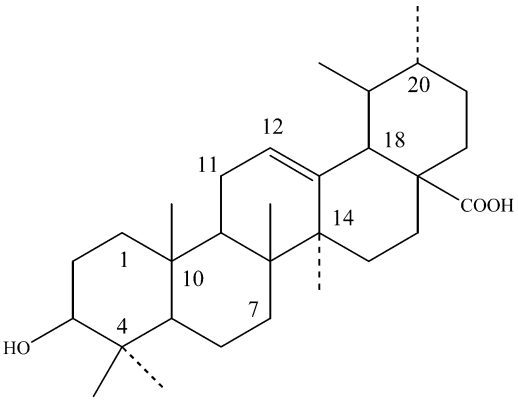
Studies concerning the metabolites present in *I. paraguariensis* differ according to the material analyzed, that is, fresh plant material or mate, the commercial product. A caffeine content of 0.9−2.2% was determined in leaves, and this content was found to depend on the age of the leaves (*6*, *7*). Clifford and Ramirez-Martinez examined five commercial samples of two types of yerba mate of South American origin, using high-performance liquid chromatography (HPLC), finding the caffeine content to be between 0.89 and 1.73% and the theobromine content to be between 0.45 and 0.88%, with very small quantities of other purines (*8*). However, the caffeine content of commercial samples can be variable because it depends greatly on the industrial treatment of the raw material (*9*). In plant material, Reginatto et al. found the methyl xanthines content of *I. paraguariensis* var. *paraguariensis* to be ∼1.8% in old leaves using HPLC analysis (*10*). The caffeine content was estimated at ∼0.65% in old leaves and at up to 1.4% in young leaves, whereas theobromine varied from 0.02% (in old leaves) to 0.27% in young leaves. Apart from caffeine analogues, caffeic acid, chlorogenic acid, and the three dicaffeoylquinic acids were found in all of the species. Among the substitutes or adulterants assayed in recent studies, *Ilex brevicuspis* showed the highest total caffeoyl derivatives content (1.9%) followed by *Ilex argentina* (0.73%) and *Ilex pseudobuxus* (0.67%) (*11*). Other chlorogenic acids that were found were ferulic acid, *p*-coumaric acid, caffeoylquinic acids, feruloylquinic acids, *p*-coumaroylqui-

* Author to whom correspondence should be addressed (e-mail verpoort@chem.leidenuniv.nl; telephone +31 71 527 4510; fax +31 71 527 4511).
† Division of Pharmacognosy, Leiden University.
§ Division of NMR, Leiden University.
# EEA Cerro Azul INTA.

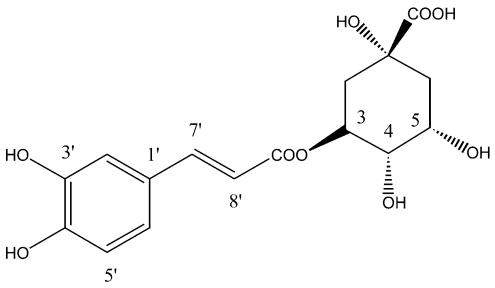
10.1021/jf0486141 CCC: $30.25
Published on Web 01/25/2005



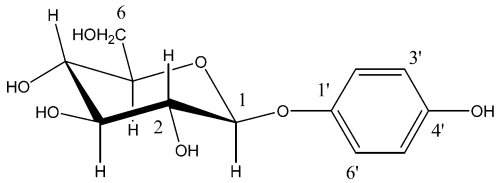
$R_1 = R_2 = R_3 = CH_3$, caffeine
$R_1 = H$, $R_2 = R_3 = CH_3$, theobromine
$R_1 = R_2 = CH_3$, $R_3 = H$, theophylline



ursolic acid



chlorogenic acid



arbutin (hydroquinone-β-glucopyranoside)

nic acids, and caffeoylferuloylquinic acids. Rutin, quercetin, and kaempferol were also found as the main flanovoids. *I. pseudobuxus* showed the highest content of rutin (0.03%), which is quite below the value obtained for *I. paraguariensis* (0.06%), whereas similar quercitin and kaempferol contents were found in the other *Ilex* species (*11*). Ursolic acid was isolated as the main triterpene, together with amyrin. The presence of triterpenoidal saponins, such as matesaponins was reported (*12*, *13*).

I. paraguariensis has many local congeneric substitutes that grow in the same habitat, the most common of which—*I. brevicuspis* as *I. argentina*, *I. brasiliensis*, *I. dumosa*, *I. integerrima*, *I. microdonta*, *I. pseudobuxus*, *I. taubertiana*, and *I. theezans* were chosen for this study. They are often used as substitutes or adulterants of *I. paraguariensis* (yerba mate) (*1*, *2*). Despite the extensive studies on the chemical composition of *Ilex* species, previous results are limited to the differentiation of particular metabolite analysis. For example, the main differences reported in previous references were a higher content of caffeoyl derivatives and flavonoids and the presence of caffeine in *I. paraguariensis* and its absence, or presence in very low amounts, in other species. However, the classification or discrimination of each *Ilex* species is still practically impossible when the classical method of analyzing a single group of metabolites is applied, and no method that could potentially contribute to their detection in mixtures has been published. For the reliable differentiation of *Ilex* species, a systematic method involving a wide variety of metabolites (metabolomic profiling) could be a very useful contribution to these issues.

The term "metabolome" has been used to describe the observable chemical profile or fingerprint of the metabolites in whole tissues (*14*). To obtain the most complete metabolomic profile, it is fundamental to use a wide spectrum analytical technique that is rapid, reproducible, and stable over time and requires only a very simple sample preparation. NMR is one of the techniques that meets those requirements. In the past decade, a number of techniques have been devised to develop NMR spectroscopy as a fingerprinting tool for the quality assessment of natural products such as food and medicinal plants. Multivariate or pattern recognition techniques such as the well-described principal component analysis (PCA) are important tools for the analysis of the data obtained by NMR. Recently, NMR in combination with PCA has been applied to the metabolomic profiling of several kinds of wines (*15*), coffees (*16*), juices (*17*), beers (*18*), and some plants (*19*, *20*).

In this study we report an NMR spectroscopic method, coupled to PCA for the metabolomic analysis of 11 *Ilex* species including two varieties. On the basis of these data, classification and discrimination are performed for the 11 *Ilex* species. This leads to a clear differentiation of *Ilex* species based on a variety of metabolites.

MATERIALS AND METHODS

Materials. Dried plant material of 11 *Ilex* species were provided by the Estación Experimental Agraria of Cerro Azul (INTA) (Misiones, Argentina). The samples were harvested 2 months prior to their use, dried for 3 min with a microwave (700 W), ground, and preserved at −18 °C. Voucher specimens are preserved in the EEA Cerro Azul. Specifications of the plant materials evaluated in this study are shown in **Table 1**. First-grade chloroform and methanol were purchased from Merck Biosolve Ltd. (Valkenswaard, The Netherlands). $CDCl_3$ (99.96%) and D_2O (99.00%) were purchased from Cambridge Isotope Laboratories Inc. (Miami, FL), and NaOD was from Cortec (Paris, France). Arbutin was obtained from Sigma (St. Louis, MO).

Table 1. *Ilex* Species Evaluated in This Study

species	region and year of seed collection (no. of voucher specimen)
I. argentina Lillo	Cerro san Javier, Tucumán, Argentina, 1991 (109) Acheral, Tucumán, Argentina, 1991 (111) Conception, Tucumán, Argentina, 1991 (112) Quebrada de San Lorenzo, Salta, Argentina, 1995 (207)
I. brasiliensis (Spreng) Loes.	Rio Branco do Sul, Paraná, Brazil, 1990 (59) Pto. Esperanza, Misiones, Argentina, 1990 (221) Reserva Biologica de Limoy, Paraguay, 1997 (226) Nueva Esperanza, Paraguay, 1997 (230)
I. brevicuspis Reissek	San Pedro, Misiones, Argentina, 1987 (4) Canoinhas, Brazil, 1989 (15) Clevelandia, Paraná, Brazil, 1991 (94) Veranopolis, Río Grande do Sul, Brazil, 1992 (119)
I. dumosa var. *dumosa* Reissek	Campo Viera, Misiones, Argentina, 1989 (7) Canoinhas, Santa Catarina, Brazil, 1989 (13) Tijucas do Sul, Paraná, Brazil, 1990 (55) Campo Bom, Río Grande do Sul, Brazil, 1992 (113)
I. dumosa var. *guaranina* Loes.	Pto. Esperanza, Misiones, Argentina, 1996 (222) Reserva Biologica de Limoy, Paraguay, 1997 (227) Hernandarias, Paraguay, 1998 (235) Hernandarias, Paraguay, 1998 (243)
I. integerrima (Vellozo) Reissek	Tijucas do Sul, Paraná, Brazil, 1990 (56) San Mateo do sul, Paraná, Brazil, 1990 (69) Iratí, Paraná, Brazil, 1990 (72) Teixeira Soares, Paraná, Brazil, 1990 (73)
I. microdonta Reissek	Sao Francisco de Paula, Río Grande do Sul, Brazil, 1992 (121) Sao Francisco de Paula, Río Grande do Sul, Brazil, 1992 (121a) Sao Francisco de Paula, Río Grande do Sul, Brazil, 1992 (121b) Parque Nacional Aparados da Serra, Río Grande do Sul, Brazil, 1992 (126)
I. paraguariensis var. *paraguariensis* St.-Hill.	Chapecó, Santa Catarina, Brazil, 1989 (28) San Antonio, Misiones, Argentina, 1989 (45) Teixeira Soares, Paraná, Brazil, 1990 (74) Ijuí, Río Grande do Sul, Brazil, 1993 (133)
I. pseudobuxus Reissek	Pontal do Sul, Paraná, Brazil, 1990 (67) Campo Bom, Río Grande do Sul, Brazil, 1992 (114) Torres, Río Grande do Sul, Brazil, 1992 (131) Tramandaí, Río Grande do Sul, Brazil, 1992 (132)
I. taubertiana Loes.	Sao Francisco de Paula, Río Grande do Sul, Brazil, 1992 (124) Sao Francisco de Paula, Río Grande do Sul, Brazil, 1992 (124a) Sao Francisco de Paula, Río Grande do Sul, Brazil, 1992 (124b) Sao Francisco de Paula, Río Grande do Sul, Brazil, 1992 (124c)
I. theezans Reissek	Major Viera, Río Grande do Sul, Brazil, 1989 (16) San Antonio, Misiones, Argentina, 1989 (46) Tijucas do Sul, Paraná, Brazil, 1990 (54) Veranopolis, Río Grande do Sul, Brazil, 1992 (118)

Methods. *Extraction.* Three hundred milligrams of ground material was placed in a centrifuge tube. Five milliliters of a 50% water/methanol mixture and 5 mL of chloroform were added to the tube followed by vortexing for 30 s and sonication for 1 min. The tube was then centrifuged at 3000 rpm for 20 min. This procedure was repeated twice. The organic fractions were transferred to a 25 mL round-bottom flask and taken to dryness with a rotary vacuum evaporator. The aqueous layer was diluted 10 times with deionized water and evaporated in a SpeedVac.

NMR Measurements. Organic fractions were dissolved in $CDCl_3$. Aqueous fractions were dissolved with D_2O to which KH_2PO_4 was added as a buffering agent. The pH of the D_2O for NMR measurements was adjusted to 6.0 using a 1 M NaOD solution. All spectra were recorded on a Bruker AV 400 NMR spectrometer operating at a proton NMR frequency of 400.13 MHz. For each sample, 128 scans were recorded with the following parameters: 0.126 Hz/point, pulse width (PW) = 4.0 μs (30°), and relaxation delay (RD) = 2.0 s. FIDs were Fourier transformed with LB = 0.3 Hz. For quantitative analysis, peak area was used. The spectra were referenced to residual solvent $CDCl_3$ at 7.26 ppm for the organic fraction and to trimethylsilanepropionic acid sodium salt (TSP) at 0.00 ppm for the aqueous fraction. Hexamethyldisilane (HMDSO, 0.01%, v/v) for $CDCl_3$ and TSP (0.01%, w/v) for aqueous fractions were used as internal standards for scaling of all NMR signals. The regions of δ 4.6−5.8 in the aqueous fraction and δ 7.28-δ 7.24 in the organic fractions were excluded from the analysis because of the residual water and $CHCl_3$ signals, respectively.

Data Analysis. The ^{1}H NMR spectra were automatically reduced to ASCII files using AMIX (v. 3.7, Bruker Biospin). Spectral intensities were scaled to HMDSO for the organic fraction and to TSP for the aqueous fraction and reduced to integrated regions of equal width (0.02

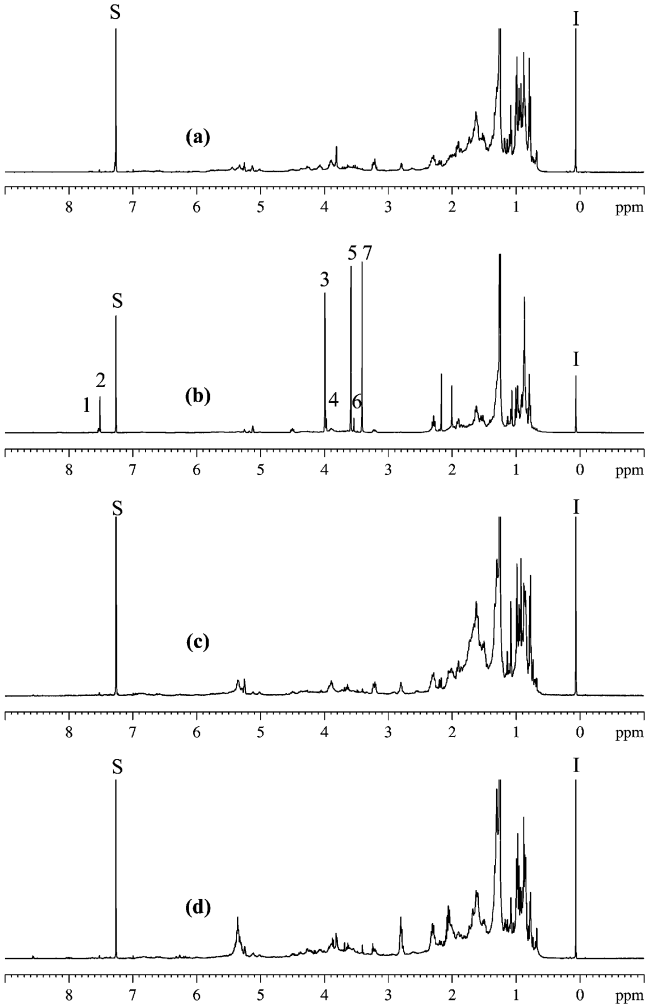


Figure 1. ^{1}H NMR spectra for organic fractions of *I. argentina* leaves (**a**), *I. paraguariensis* leaves (**b**), *I. pseudobuxus* leaves (**c**), and *I. tauberiana* leaves (**d**). Peaks: 1, H-8 of theobromine; 2, H-8 of caffeine; 3, 7-methyl of caffeine; 4, 7-methyl of theobromine; 5, 3-methyl of caffeine; 6, 3-methyl of theobromine; 7, 1-methyl of caffeine; S, residual solvent signal of $CDCl_3$; I, internal standard of HMDSO.

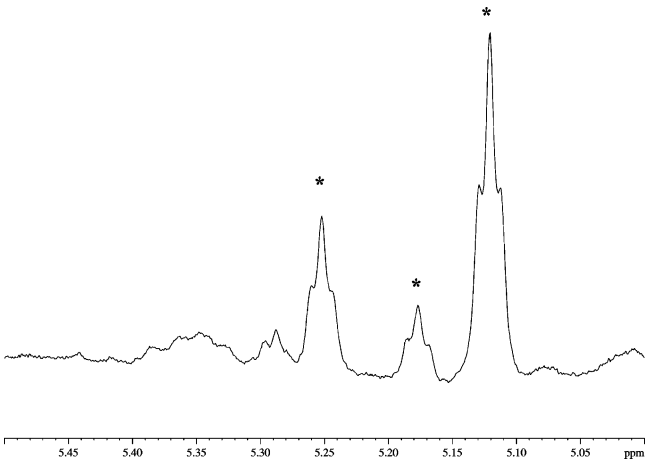


Figure 2. ^{1}H NMR spectra for organic fraction of *I. paraguariensis* leaves in the range of δ 5.0–5.5. *, possible signals of H-12 protons of ursolic acid analogues.

ppm) corresponding to the region of δ 0.40−10.00. PCA and discriminant analysis were performed with the SIMCA-P 10.0 software (Umetrics, Umeå, Sweden).

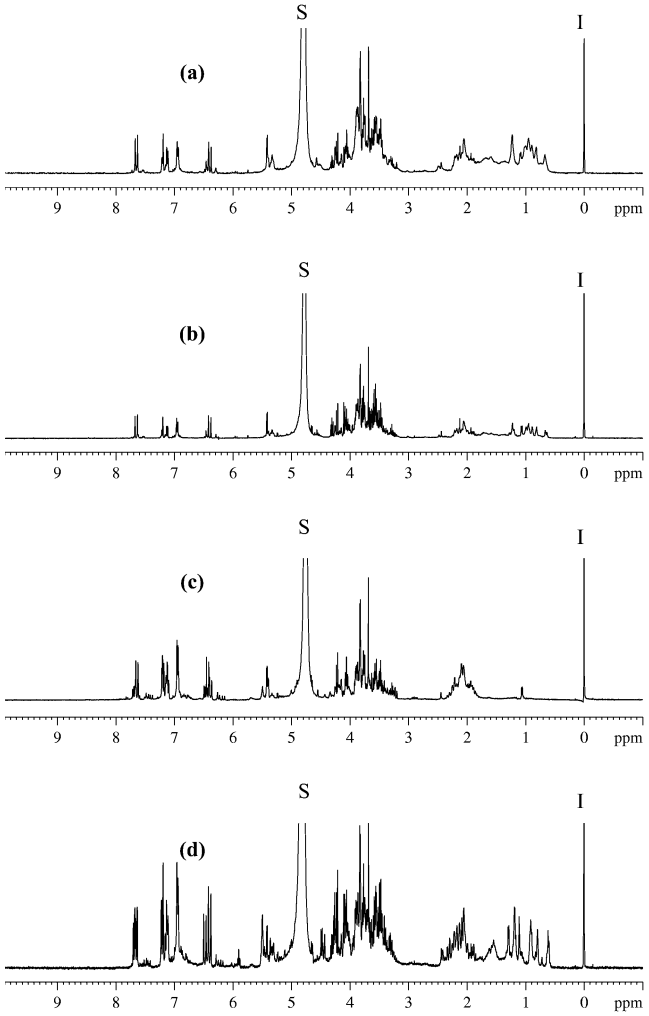


Figure 3. ^{1}H NMR spectra for aqueous fractions of *I. dumosa* var. *dumosa* leaves (**a**), *I. dumosa* var. *guaraniana* leaves (**b**), *I. paraguariensis* leaves (**c**), and *I. pseudobuxus* leaves (**d**). Peaks: S, residual solvent signal of HDO; I, internal standard of TSP.

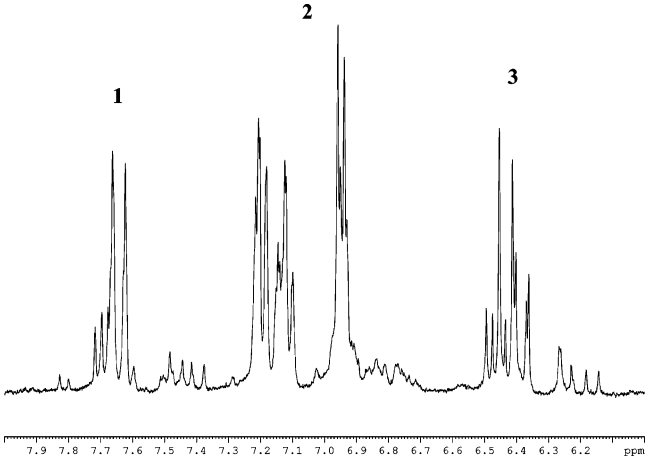


Figure 4. ^{1}H NMR spectra for aqueous fraction of *I. paraguariensis* leaves in the range of δ 6.0–8.0. Peaks: 1, H-7′ region of phenylpropanoids; 2, aromatic region of phenylpropanoids; 3, H-8′ region of phenylpropanoid.

HPLC Analysis. A Waters HPLC system equipped with a 626 pump, a 2996 photodiode array detector fixed at 282 nm, and a 717 plus autosampler (Waters, Milford, MA) was used for arbutin determination. Twenty microliter samples were injected onto an Inertsil ODS-3 (4.6 × 250 mm, *s*-3 μm) (GL Sciences Inc., Tokyo, Japan) column and eluted with a linear gradient starting at a proportion of 90:10 of 1%

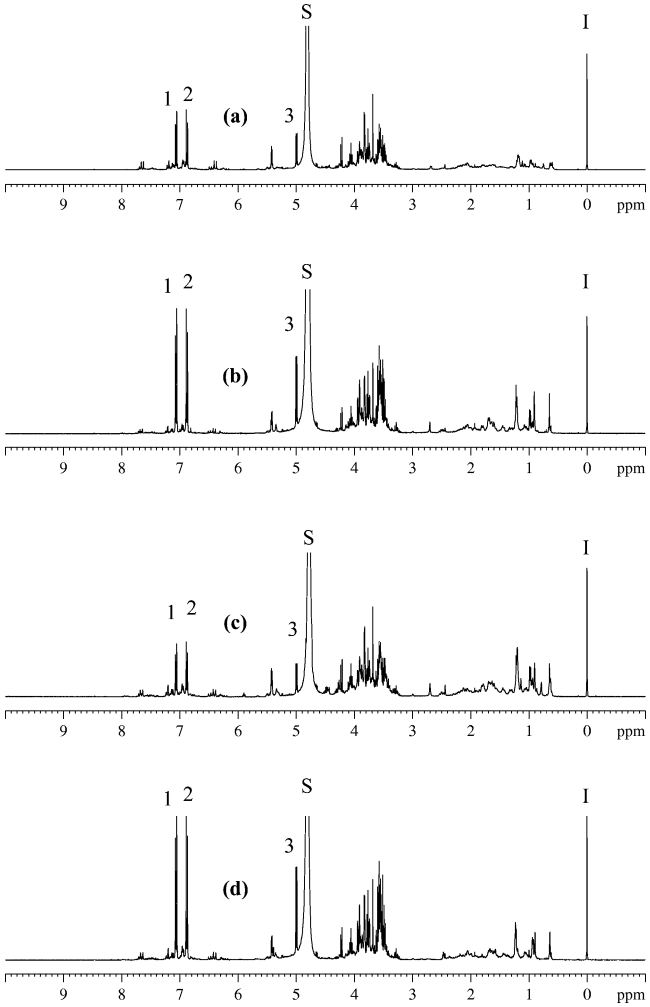


Figure 5. ^{1}H NMR spectra for aqueous fractions of *I. argentina* leaves (**a**), *I. brasiliensis* leaves (**b**), *I. brevicuspis* leaves (**c**), and *I. theezans* leaves (**d**). Peaks: S, residual solvent signal of HDO; I, internal standard of TSP; 1, H-2′ and H-6′ of arbutin; 2, H-3′ and H-5′ of arbutin; 3, H-1 of arbutin.

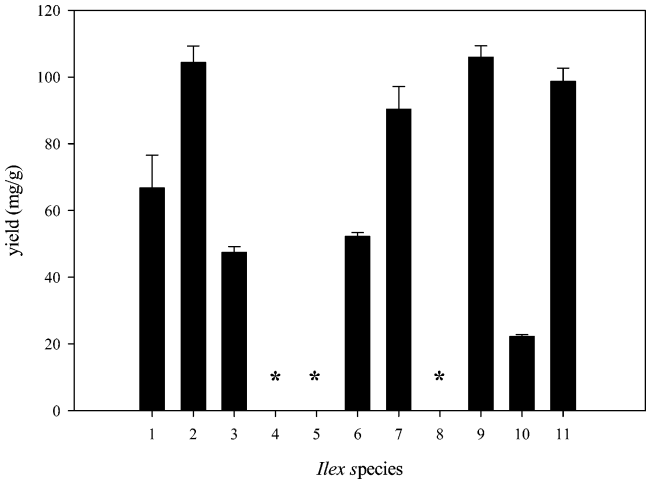


Figure 6. Yield (mg/g) of arbutin in *Ilex* species obtained from HPLC analysis: 1, *I. argentina*; 2, *I. brasiliensis*; 3, *I. brevicuspis*; 4, *I. dumosa* var. *dumosa*; 5, *I. dumosa* var. *guaranina*; 6, *I. integerrima*; 7, *I. microdonta*; 8, *I. paraguariensis* var. *paraguariensis*; 9, *I. pseudobuxus*; 10, *I. taubertiana*; 11, *I. theezans*; *, not detected. Results are based on triplicate analysis.

AcOH/H_2O/1% AcOH/MeOH for 4.5 min and then changing to 70:30 in 26 min and to 90:10 in 20 min. The flow rate was 1.0 mL/min.

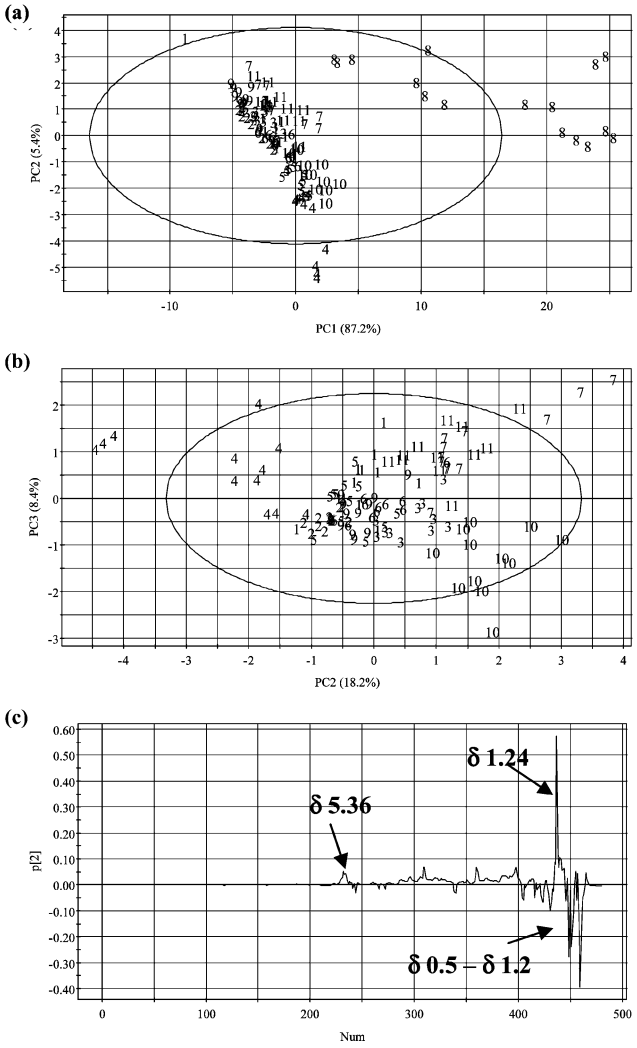


Figure 7. Score plot of discriminating PC scores of organic fraction for *Ilex* species with *I. paraguariensis* (**a**) and without *I. paraguariensis* (**b**) following PCA analysis: 1, *I. argentina*; 2, *I. brasiliensis*; 3, *I. brevicuspis*; 4, *I. dumosa* var. *dumosa*; 5, *I. dumosa* var. *guaranina*; 6, *I. integerrima*; 7, *I. microdonta*; 8, *I. paraguariensis* var. *paraguariensis*; 9, *I. pseudobuxus*; 10, *I. taubertiana*; 11, *I. theezans*.

RESULTS AND DISCUSSION

Visual Inspection of ^{1}H NMR Spectra and Assignments of the Compounds. Among the *Ilex* species evaluated in this study, only *I. paraguariensis* was found to contain caffeine and theobromine, whereas theophylline was not detected. This finding is in accordance with reports of caffeine and theobromine in *I. paraguariensis* leaves dating back to the 19th century, whereas the presence of theophylline, reported in very small quantities (*21*, *22*), is a matter of controversy, as other researchers did not detect this substance (*8*, *23*, *24*). Characteristic signals due to one purine proton at δ 7.51 (1H, s) and three *N*-methyls at δ 3.99, 3.59, and 3.41 (3H each, s) in the ^{1}H NMR spectrum of organic fractions of *I. paraguariensis* are in accordance with caffeine. In addition to these caffeine signals, δ 7.53 (1H, s), 3.97 (3H, s), and 3.54 (3H, s) were assigned to theobromine (**Figure 1**). Theophylline was not detected in *I. paraguariensis* leaves. The main triterpenes of *Ilex* species were found to be ursolic acid analogues (*12*, *13*). In the region of δ 5.0−5.5, characteristic olefinic protons (H-12) of triterpenes are detected at δ 5.12 (t, J = 3.3 Hz), 5.17 (t, J = 3.3 Hz), and 5.29 (t, J = 3.3 Hz) (**Figure 2**). Methyl signals of triterpenes

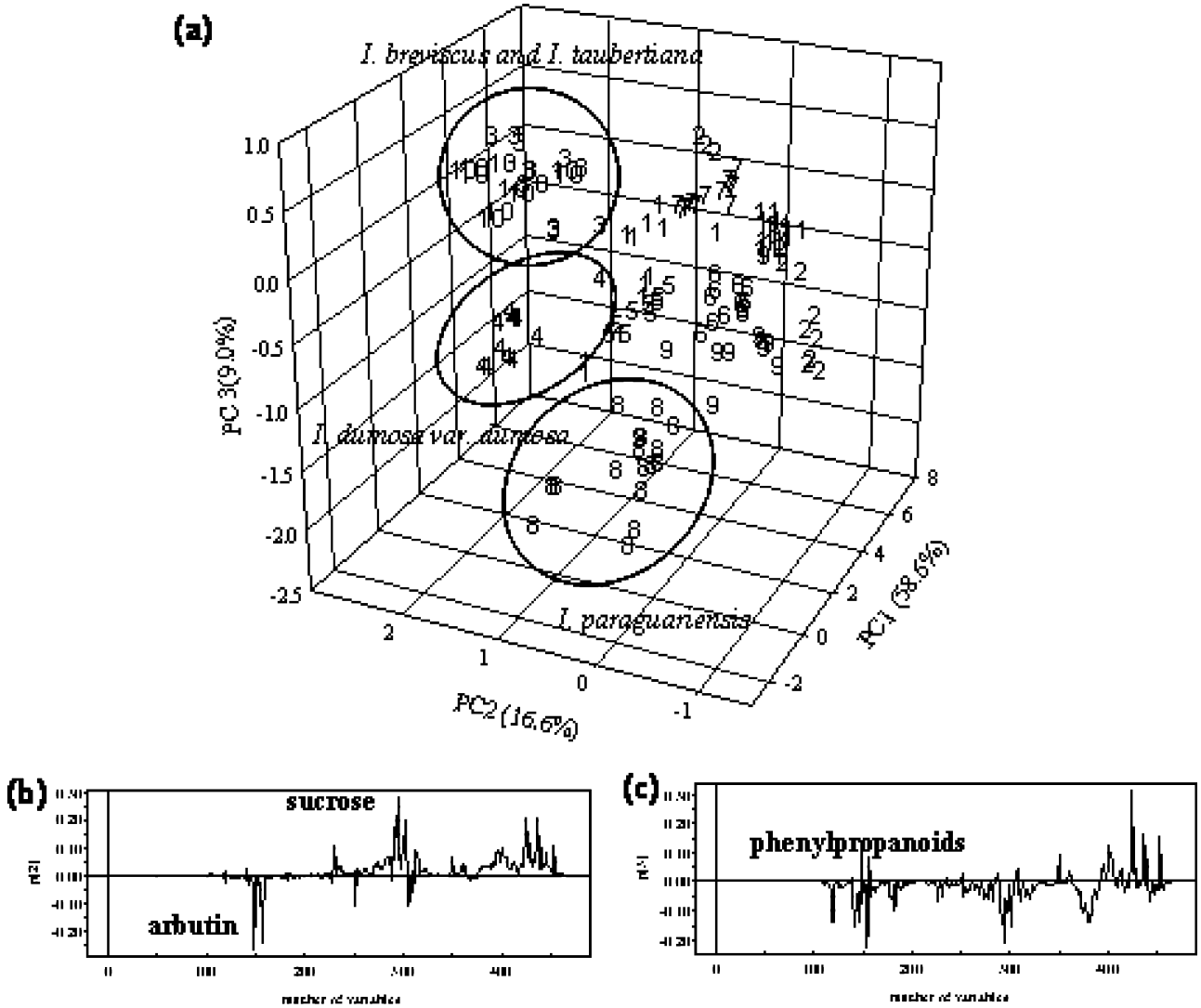


Figure 8. Score plot of PC1, PC2, and PC3 scores of aqueous fraction for *Ilex* species following PCA (**a**) and loading plots for PC2 (**b**) and PC3 (**c**): 1, *I. argentina*; 2, *I. brasiliensis*; 3, *I. brevicuspis*; 4, *I. dumosa* var. *dumosa*; 5, *I. dumosa* var. *guaranina*; 6, *I. integerrima*; 7, *I. microdonta*; 8, *I. paraguariensis* var. *paraguariensis*; 9, *I. pseudobuxus*; 10, *I. taubertiana*; 11, *I. theezans*.

in the range of δ 0.7−1.2 were also found to be discriminating ^{1}H NMR signals of *Ilex* species.

The ^{1}H NMR spectra of the aqueous fractions for the *Ilex* species are shown in **Figure 3**. The patterns in the aromatic region (δ 6.0−8.0) are quite different from each other. Previous reports have shown that some phenylpropanoids such as caffeic acid, chlorogenic acid, 3,4-dicaffeoylquinic acid, 3,5-dicaffeoylquinic acid, and 4,5-dicaffeoylquinic acid are the main characteristic metabolites of *Ilex* species (*11*). In accordance with this paper, the signals of the main differentiating aromatic compounds in the extracts were assigned to phenylpropanoids (**Figure 3**). The ^{1}H NMR spectrum of *I. paraguariensis* (**Figure 4**) is in accordance with the spectrum of phenylpropanoids, showing the typical signals due to two trans olefinic protons (J = 15−16 Hz) in the region of δ 6.1−6.5 (H-8′) and δ 7.6−7.7 (H-7′). The complex pattern of the ^{1}H NMR spectrum in δ 6.9−7.3 shows that this plant contains several phenylpropanoids such as caffeic acid, *p*-coumaric acid, and their glycosides.

Intriguingly, intense signals at δ 7.06 (1H, d, J = 9.0 Hz), 6.88 (1H, d, J = 9.0 Hz), and 4.89 (1H, d, J = 7.6 Hz) appeared in *I. argentina*, *I. brasiliensis*, *I. integerrima*, *I. microdonta*, *I. taubertiana*, and *I. theezans* (**Figure 5**). These signals were assigned to H-2′, H-3′, and the anomeric proton of glucose in arbutin, respectively. This was confirmed by 2D-NMR spectra such as ^{1}H−^{1}H COSY, HMQC, and HMBC and comparison with the reference compound arbutin. There have been no previous reports of arbutin in *Ilex* species. In this study, however, it was found that some *Ilex* species contain arbutin as a major metabolite. For further confirmation and quantitative analysis of arbutin in *Ilex* species evaluated in this study, HPLC analysis was performed. Among the species studied, *I. pseudobuxus* showed the highest content of arbutin (106.0 mg/g) (**Figure 6**).

PCA. PCA is an unsupervised clustering method requiring no knowledge of the data set and acts to reduce the dimensionality of multivariate data while preserving most of the variance within it (*25*). The data for PCA can be scaled in different ways. If the data are mean-centered with no scaling, then a covariance matrix is produced, but if the data are mean-centered and the columns of the data matrix are scaled to unit variance, a correlation matrix is produced (*26*). Both of the methods were applied to the ^{1}H NMR data set of *Ilex* species, the covariance method showing a better separation of the species. Application of PCA to the organic fractions, resulted in a good separation of *I. paraguariensis* from other species by PC1 (**Figure 7a**).

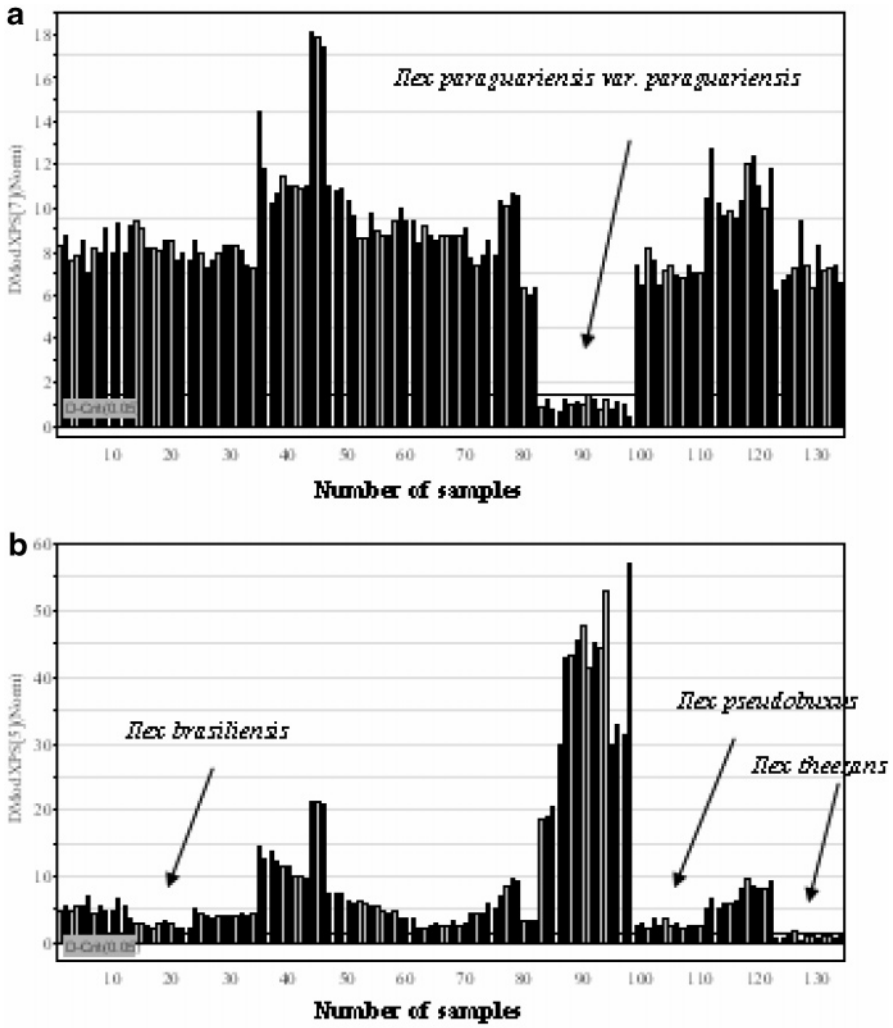


Figure 9. DModX plot for *Ilex* species prediction set of organic fractions of *I. paraguariensis* var. *paraguariensiss* using the first seven PCs (**a**) and of *I. theezans* using the first five PCs (**b**): 1–12, *I. argentina*; 13–23, *I. brasiliensis*; 24–34, *I. brevicuspis*; 35–46, *I. dumosa* var. *dumosa*; 47–58, *I. dumosa* var. *guaranina*; 59–70, *I. integerrima*; 71–82, *I. microdonta*; 83–98, *I. paraguariensis* var. *paraguariensis*; 99–110, *I. pseudobuxus*; 111–122, *I. taubertiana*; 123–134, *I. theezans*.

This separation is due to the signals of caffeine, theobromine, and triterpenoids. For the detailed inspection of other *Ilex* species, PCA was done for all of the species excluding *I. paraguariensis*. As shown in **Figure 7b**, *I. dumosa* var. *dumosa* and *I. tauberiana* were clearly separated from other species. *I. dumosa* var. *dumosa* contains more triterpenoids and *I. tauberiana* has a higher amount of fatty acids, as can be concluded from the loading plot of PC2, in which a higher PC2 value indicates higher signals at δ 5.36 (olefinic) and 1.24 (methylene) of fatty materials and lower terpenoidal signals in the range of δ 0.5−1.2 (**Figure 7c**).

As a next step, the focus was placed on the results obtained from the aqueous fraction. For a clear separation, however, PC3 is needed apart from PC1 and PC2. As seen in **Figure 8a**, there is a clear discrimination between *I. paraguariensis* and other *Ilex* species. Notably, aqueous fractions of *I. dumosa* var. *dumosa*, *I. breviscus*, and *I. taubertiana* show unique metabolomic fingerprints. This separation took place in the first three principal components, which cumulatively accounted for 84.2% of the variation. The separation between *I. paraguariensis* and other species was easily achieved with the PC2 and PC3 values. *I. paraguariensis* has a lower PC2 and PC3. The PC2 value is affected by the amount of arbutin and sucrose, and a higher PC2 value means lower phenylpropanoids and sucrose content (**Figure 8b**). PC3 was largely influenced by the quantity of phenylpropanoids, and its lower value indicated a higher amount of phenylpropanoids (**Figure 8c**).

Classification of *Ilex* Species Based on PCA. Pattern recognition is often described as a procedure for formulating rules of classification (*27*). The most encountered goal of a pattern recognition application is classification (*28*). Using a collection of knowns and a classification rule, a set of unknowns is classified. On the basis of given classes, each of which contains a number of observations mapped by a multitude of variables, guidelines and rules are developed that make it possible to classify new observations as similar or dissimilar to the members of the existing classes (*25*). Data that was observed in a class of one *Ilex* species was classified by soft independent modeling of class analogy (SIMCA) (*29*). Using this method, each class of *Ilex* species was modeled separately by disjoint PC models. On the basis of the residual variation of each class, the distance to the model (DModX) of each observation was computed.

Prior to the classification, PCA was carried out for each species to obtain optimum PCs, which were necessary to build the model of each species. After each particular model based

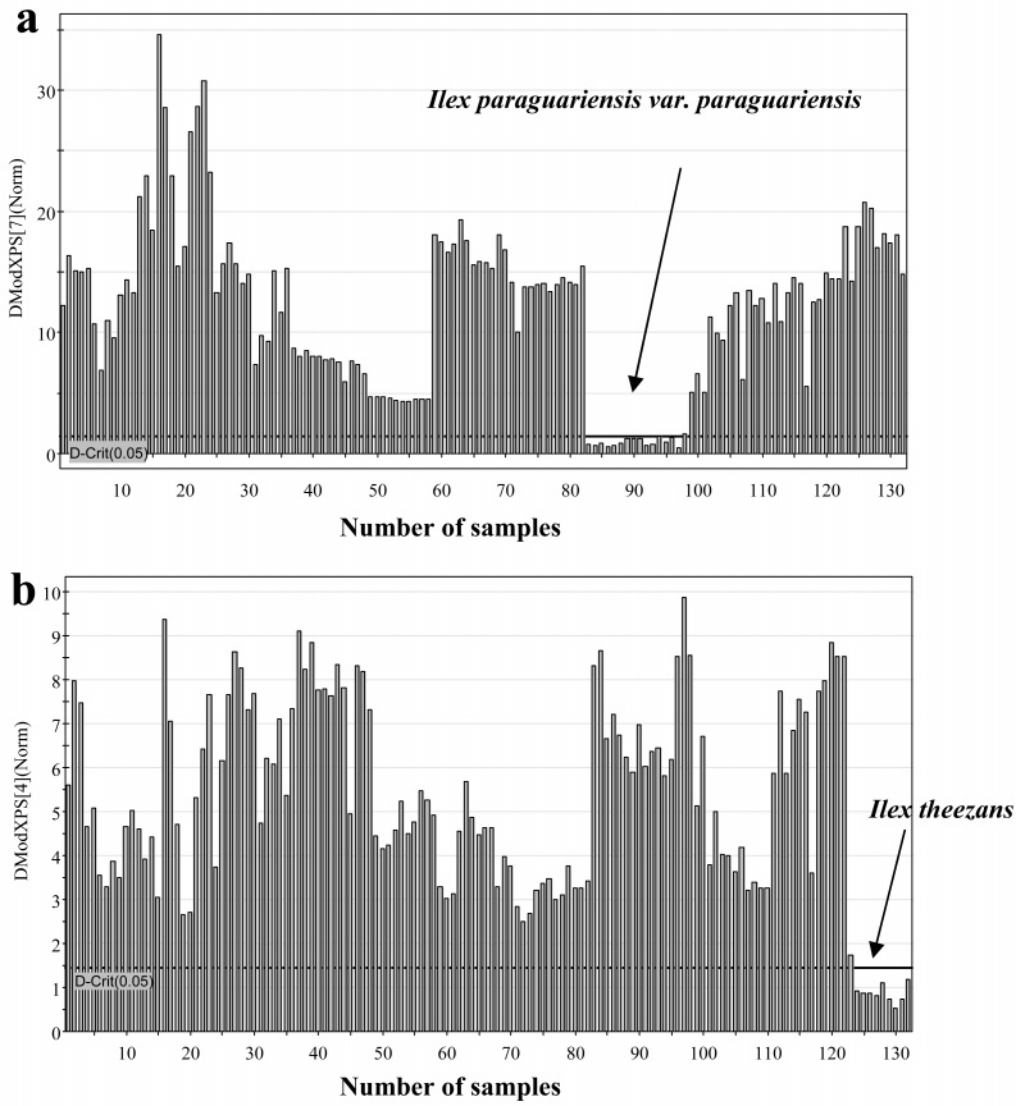


Figure 10. DModX plot for *Ilex* species prediction set of aqueous fractions of *I. paraguariensis* var. *paraguariensiss* using the first seven PCs (**a**) and of *I. theezans* using the first four PCs (**b**): 1–12, *I. argentina*; 13–24, *I. brasiliensis*; 25–36, *I. brevicuspis*; 37–48, *I. dumosa* var. *dumosa*; 49–58, *I. dumosa* var. *guaranina*; 59–70, *I. integerrima*; 71–82, *I. microdonta*; 83–98, *I. paraguariensis* var. *paraguariensis*; 99–110, *I. pseudobuxus*; 111–122, *I. taubertiana*; 123–132, *I. theezans*.

on optimum PCs (e.g., the first seven PCs were used for the metabolites obtained from organic fractions of *I. paraguariensis*) had been constructed, all species were compared using the distance to each species model. In the classification based on the metabolites obtained from organic fractions, most of the species do not overlap and showed unique metabolomic profiles as an example of *I. paraguariensis* (**Figure 9a**) with the exception of *I. pseudobuxus*, *I. brasiliensis*, and *I. theezans* (**Figure 9b**). In the classification of metabolites obtained from aqueous fractions, no species overlapped. As shown in the case of *I. paraguariensis* in **Figure 10a**, even the species *I. pseudobuxus*, *I. brasiliensis*, and *I. theezans*, the metabolite profiles of which could not completely be resolved for the organic fractions, were clearly distinguished from each other with the metabolomic profile of the aqueous fractions (**Figure 10b**).

This study proves that it is possible to discriminate the 11 *Ilex* species, *I. argentina*, *I. brasiliensis*, *I. brevicuspis*, *I. dumosa* var. *dumosa*, *I. dumosa* var. *guaranina*, *I. integerrima*, *I. microdonta*, *I. paraguariensis* var. *paraguariensis*, *I. pseudobuxus*, *I. taubertiana*, and *I. theezans*, by multivariate analysis of their metabolite fingerprints obtained by ^{1}H NMR spectra of crude extracts of the plant materials. The major compounds contributing to the discrimination were found to be several phenylpropanoids and arbutin. In particular, arbutin, which has not been reported yet as a component of *Ilex* species, was found to be a discriminating metabolite (biomarker) for *I. argentina*, *I. brasiliensis*, *I. brevicuspis*, *I. integerrima*, *I. microdonta*, *I. pseudobuxus*, *I. taubertiana*, and *I. theezans*. However, arbutin was not present in a detectable amount in the samples of *I. dumosa* var. *dumosa*, *I. dumosa* var. *guaranina*, and *I. paraguariensis* var. *paraguariensis* that were analyzed. Besides these major phenolic metabolites, a number of putative minor metabolites also play a role in differentiating *Ilex* species.

Alongside simple PCA used for the reduction of the NMR data set obtained from the metabolites, classification based on each species model was done and obviously complements the differentiation of *Ilex* species based on the metabolite profiles.

The method using ^{1}H NMR and multivariate analysis may afford consistent discrimination of *Ilex* species based on

metabolomic profiling as a tool for chemotaxonomical studies. This method could also be useful for quality control issues, which are critical to the yerba mate producing industry, such as detection of adulterants, leaf/stalk content, detection of inadequate drying or conservation of plants, and standardization of blends among others for which there is not an adequate analytical solution at this time.

LITERATURE CITED

(1) Graham H. N. Mate. *Prog. Clin. Biol. Res.* **1984**, *58*, 179−183.

(2) Giberti, G. Los parientes silvestres de la yerba mate y el problema de su adulteración. *Dominguezia* **1989**, *7*, 1−22.

(3) Gonzales, A.; Ferreira, F.; Vázques, A.; Moyna, P.; Alonso Paz, E. Biological screening of Uruguayan medicinal plants, *J. Ethnopharmacol*. **1993**, *39*, 217−220.

(4) Guliucci, A. Antioxidant effects of *Ilex paraguariensis*: induction of decreased oxidability of human LDL in vivo. *Biochem. Biophys. Res. Commun.* **1996**, *224*, 338−344.

(5) Gorzalczany, S.; Filip, R.; Alonso, M. R.; Mino, J.; Ferraro, G. E.; Acevedo, C. Choleretic effect and intestinal populsion of mate (*Ilex paraguariensis*) and its substitutes or adulterants. *J. Ethnopharmacol*. **2001**, *75*, 291−294.

(6) Graham, H. Mate. In *The Methylxanthine Beverages and Foods*: *Chemistry*, *Consumption and Health Effects*; Spiller, G. A., Ed.; Liss: New York, 1984; pp 179−183.

(7) Bertoni, M. H.; Prat Kircun, S. D.; Kanzig, R. G.; Cattaneo, Y. P.; Hojas frescas de *Ilex paraguariensis* St.-Hill. III. *An. Asoc. Quim. Argent.* **1992**, *80*, 493−501.

(8) Clifford, M. N.; Ramirez-Martinez, J. R. Chlorogenic acids and purine alkaloids contents of maté (*Ilex paraguariensis*) leaf and beverage. *Food Chem.* **1990**, *35*, 13−21.

(9) Schmalko, M. E.; Alzadora S. M. Color, chlorophyll, caffeine and water content variation during yerba mate processing. *Drying-Technol* **2001**, *19*, 599−610.

(10) Reginatto, F. H.; Athayde, M. L.; Gosmann, G.; Schenkel, E. P. Methylxanthines accumulation in *Ilex* species—caffeine and theobromine in erva—mate (*Ilex paraguariensis*) and other *Ilex* species. *J. Braz. Chem. Soc*. **1999**, *10*, 443−446.

(11) Filip, R.; Lopez, P.; Giberti G.; Coussio, J.; Ferraro, G. Phenolic compounds in seven south American *Ilex* species. *Fitoterapia* **2001**, *72*, 774−778.

(12) Kraemer, K. H.; Taketa, A. T. C.; Schenkel, E. P.; Gosmann, G.; Guillaume, D. Matesaponin 5, a highly polar saponin from *Ilex paraguariensis*. *Phytochemistry* **1996**, *42*, 1119−1122.

(13) Taketa, A. T. C.; Schmittmann-Schlager, T.; Guilaume, D.; Gosmann, G.; Schenkel, E. P. Triterpenoid glycosides and a triterpene from *Ilex brevicuspis*. *Phytochemistry* **2000**, *53*, 901−904.

(14) Ott, K.-H.; Araníbar, N.; Singh, B.; Stockton, G. W. Metabonomics classifies pathways affected by bioactive compounds. Artificial neural network classification of NMR spectra. *Phytochemistry* **2003**, *62*, 971−985.

(15) Brescia, M. A.; Caldarola, V.; De Giglio, A.; Benedetti, D.; Fanizzi, F. P.; Sacco, A. Characterization of the geographical origin of Italian red wine based on traditional and nuclear magnetic resonance spectrometric determinations. *Anal. Chim. Acta* **2002**, *458*, 117−186.

(16) Charlton, A. J.; Farrington, W. H. H.; Brereton, P. Application of ^{1}H NMR and multivariate statistics for screening complex mixtures: quality control and authenticity of instant coffee. *J. Agric. Food Chem.* **2002**, *50*, 3098−3103.

(17) Vogels, J. T. W. E.; Terwel, L.; Tas, A. C.; Van den Berg, F.; Dukel, F., Van der Greef, J. Detection of adulteration in orange juices by a new screening method using proton NMR spectroscopy in combination with pattern recognition techniques. *J. Agric. Food Chem.* **1996**, *44*, 175−180.

(18) Duarte, I.; Barros, A.; Belton, P. S.; Righelato, R.; Spraul, M.; Humpfer, E.; Gil, A. M. High-resolution nuclear magnetic resonance spectroscopy and multivariate analysis for the characterization of beer. *J. Agric. Food Chem.* **2002**, *50*, 2475−2481.

(19) Forveille, L.; Vercauteren J.; Rutledge, D. N. Multivariate statistical analysis of two-dimensional NMR data to differentiate grapevine cultivars and clones. *Food Chem.* **1996**, *57*, 441−450.

(20) Nord, L. I.; Kenne, L.; Jacobsson, S. P. Multivariate analysis of ^{1}H NMR spectra for saponins from *Quilaja saponaria* Molina. *Anal. Chim. Acta* **2001**, *446*, 199−209.

(21) Mazzafera, P. Caffeine, theobromine and theophylline distribution in *Ilex paraguariensis*. *Rev. Brasileira Fisiol. Vegetal* **1994**, *6*, 49−151.

(22) Vázquez, A.; Moyna, P. Studies on mate drinking. *J. Ethnopharmacol.* **1986**, *18*, 267−272.

(23) Ashihara, H. Purine metabolism and the biosynthesis of caffeine in maté leaves. *Phytochemistry* **1993**, *33*, 1427−1430.

(24) Filip, R.; Lopez, P.; Coussio, J.; Ferrara, G. Mate substitutes or adulterants: study of xanthine content. *Phytother. Res.* **1998**, *12*, 129−131.

(25) Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Multi- and Megavariate Data Analysis*; Umetrics Academy: Umeå, Sweden, 2001; pp 43−70.

(26) Ward, J.; Harris, C.; Lewis, J.; Beale, M. H. Assessment of ^{1}H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. *Phytochemistry* **2003**, *62*, 949−957.

(27) Albano, C.; Dunn, W. J., III; Edlund, U.; Johansson, E.; Nordén, B.; Sjöström, M.; Wold, S. Four levels of pattern recognition, *Anal. Chim. Acta* **1978**, *103*, 429−443.

(28) Kowalski, B. R.; Bender, C. F. Pattern recognition. A powerful approach to interpreting chemical data. *J. Am. Chem. Soc.* **1972**, *94*, 5632−5639.

(29) Wold, S.; Albano, C.; Dunn, W. J., III; Edlund, U.; Esbensen, K.; Geladi, P.; Helberg, S.; Johansson, E.; Lindberg, W.; Sjöström, M. Multivariate data analysis in chemistry. In *Chemometrics*: *Mathematics and Statistics in Chemistry*; Kowalski, B. R., Ed.; Reidel Publishing: Dordrecht, The Netherlands, 1984.

Received for review August 19, 2004. Revised manuscript received November 30, 2004. Accepted December 3, 2004. The financial support of the van Leersumfonds (KNAW) is gratefully acknowledged.

JF0486141