RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

A Targeted Metabolomics Approach to Characterize Acetogenin Profiles in Avocado Fruit (*Persea americana* Mill.)

Rodríguez-López, Carlos Eduardo; Hernández Brenes, Carmen; and Díaz de la Garza, Rocío I.*

Tecnologico de Monterrey, Campus Monterrey, Eugenio Garza Sada 2501, Monterrey, N.L., México, 64849.

* Corresponding author: rociodiaz@itesm.mx; (+52) 818 328 4262

Abstract:

Lauraceous acetogenins are fatty acid derivatives with an odd-carbon aliphatic chain found in avocado (*Persea americana* Mill.). These compounds display a wide range of bioactivities that makes them candidates for use as antimicrobial and proapoptotic agents in food industry and against cancer cells respectively. Existent knowledge about its metabolism *in planta* is scarce. This work quantifies eight different acetogenins accumulated in fruit tissues (peel, seed, and pulp) from 22 avocado cultivars to sample the existing variation using a targeted metabolomics approach. Multivariate analyses uncovered correlations among acetogenins present in fruit tissues and their chemical backbone that allowed a proposal for classifying them in three families (Avocatins, Pahuatins and Persenins). Seed acetogenin profile differed from that of pulp and peel, which while different in concentration (peel accumulated low acetogenin amounts), had the same profile. Acetogenins from samples of known origin were also separated by variety using descriptive Linear Discriminant Analysis (LDA), and a chemotaxonomic model was generated via predictive LDA and was tested on samples from unknown origin. This work effectively sampled acetogenin contents and profile variability in seed (1.09 – 8.33 mg/g FW), peel (0.22 – 12.5 mg/g FW), and pulp (0.49 – 9.58 mg/g FW) from avocado fruit, as well as provides a putative classification to seven avocado cultivars. Results from this work show that the eight acetogenins followed are produced in all 22 avocado cultivars, which points to conserved metabolism among avocado plants.

Keywords: Lauraceous acetogenins, aliphatic acetogenins, avocatins, chemotaxonomy, targeted metabolomics, avocado screening, Persin, Persenone A, Antifungal Diene

Introduction

Avocado (*Persea americana* Mill.) fruit is one of the most nutrient-dense fruits available, it contains significant amounts of 12 (out of 13) vitamins, and it is an abundant source of vegetable oil rich in unsaturated fatty acids, accounting for 15-30% fresh weight and almost the total energy content of the fruit.[1] Apart from the nutritional qualities of the fruit, avocado has been used as a medicinal plant; particularly leaf extracts are used for treating a variety of illnesses in indigenous medicine, ranging from hypertension to diabetes.[2]

One of the main active components detected in avocado leaves and fruits are lauraceous acetogenins, which are fatty acid derivatives that typically contain an odd-carbon aliphatic chain (17, 19 or 21) and an acetoxy group that contributes two additional carbons.[3] Acetogenins bioactivities have been studied and they have a broad activity range that includes antimicrobial,[4] antifungal,[5] inhibition of the production of nitric oxide and superoxide in cells,[6, 7] selective pro-apoptotic activity against several cancer cell lines,[2,8-9] and recently, promising activity against Acute Myeloid Leukemia (AML) cell lines.[10] Bactericidal and sporostatic capacities have specifically increased the interest of food industry, due to their potential use as food additives.[11] Their lipophilic properties, as well as the increasing demand for additives from natural origin and the fact that they are already being consumed by humans at bioactive levels, makes their potential uses quite promising.[12]

As an approach for studying acetogenin functions *in planta*, several works have attempted to correlate their concentrations with particular resistance traits, effectively demonstrating toxicity of an acetogenin-rich extract against late instars of *Spodoptera exigua*.[13] Persin ((Z,Z)-1-acetoxy-2-hydroxy-12,15-heneicosadien-4-one) accumulation was followed as a response to *Colletotrichum gloeosporiodes* infection in different avocado cultivars;[14] although no correlation was found between resistant variants and Persin concentration in leaves,[15] other acetogenins not considered in that work

may have contributed to the trait. It has been observed that acetogenin bioactivity is highly dependent on the aliphatic chain structure, to such an extent that a change from an olefinic (1,2,4-trihydroxy heptadec-16-ene; avocadene) to an ethyne (1,2,4-trihydroxy heptadec-16-yne; avocadyne) bond enhances the pro-apoptotic effects against a human prostate adenocarcinoma cell line (PC-3) more than 7-fold.[8]

Despite the impact of chemical differences on the bioactivity of these moieties, prior works have mainly focused on measuring a single acetogenin, mainly Persenone A or Persin, in avocado tissues. Moreover, the majority of existing literature reports the analysis of only three avocado cultivars that include 'Reed' (*P. americana* var. *guatemalensis*),[16] 'Fuerte' (*P. americana* var. *drymifolia*),[16, 17] and 'Hass' (*P. americana* var. *guatemalensis* × *drymifolia*).[17] To the best of our knowledge, only a single work has evaluated Persin levels in avocado leaves from 21 different avocado cultivars and strains,[15] and no work has been published concerning other tissues (particularly fruit) or different acetogenin derivatives.

Thus, the objective of this study was to characterize and quantify acetogenins in fruits from 22 avocado cultivars in order to sample the existing natural variation. The present work also was undertaken to assess acetogenin variations in fruit tissues (peel, seed and pulp), considering the genotypic diversity of these cultivars. A targeted metabolomics approach that involved uni- and multivariate techniques was used, along with classification algorithms, in the pursuit of establishing possible chemotaxonomic rules distinguishing the cultivars group of origin. This is the first study which, taking advantage of a targeted metabolomics approximation, focuses on the characterization of avocado fruit acetogenins.

Experimental

Plant Material

Different avocado (*P. americana*) cultivars, shown in Supplementary Figure 1, were sampled from the Fundación Sánchez Colín – CICTAMEX collection on October the 6th, 2011. Fruits from each cultivar were harvested at mature green stage from single trees grown in the same orchard, thus subjected to the same growing conditions. CICTAMEX's experimental field is located in Coatepec Harinas, Estado de México, México (18° 55' N, 99° 45' W; 2240 m Above Mean Sea Level (AMSL)). This study included a total of 22 cultivars, 19 cultivars from the CICTAMEX collection, 1 sample of the commercially used 'Hass' cultivar, and 2 other uncharacterized landraces; which were grown in "Los Catorce" (2000 m AMSL), near Matehuala, San Luis Potosí, México (23º 39' N, 100º 39' W; 1850 m AMSL) and were kindly donated by Gerónimo Cano. These landraces were also collected on 6th of October, 2011 and labeled accordingly: "Los Catorce Chico" (L14Ch), and "Los Catorce No Endémico" (L14NE). Samples were air shipped in closed containers with activated charcoal, and upon arrival all fruits were stored at 4° C for two days, following freezing at –20° C for another 48 hours, before being stored at -80 °C for further analysis. 'Hass' avocados, donated by H.E.B. México, were collected as soon as they arrived from a commercial orchard in Michoacán, México (near Uruapan, 19°25 N, 102°03W; 1620 m AMSL) and followed the same freezing protocol mentioned above.

Acetogenin Extraction

Extractions were made as previously reported by Rodríguez-Sánchez *et al*.[18] with modifications based on a purification procedure by Prusky.[17] Briefly, tissues were separated from frozen fruits in the following manner: peels (1g) were separated using a razor blade; pulp (2g) was cut in even, longitudinal slices; and cotyledons (2g) were fragmented while frozen with the help of a hammer, mixed and then weighted. Acetogenin extraction was achieved by adding 15 mL of acetone, followed

by homogenization with a Polytron (Ultra-Turrax T25, IKA-Werke, Germany) at 16,000 rpm for 3 min and placement in an ultrasonic bath (Model 50T, VWR, Pennsylvania, U.S.A.) for 1 min. Extracts were then shaken in a multi-shaker (Lab-line Incubator-Shaker, India) at room temperature for 15 minutes, and then clarified by centrifugation at 10,000 g for 10 min. From this extract, a 1 mL aliquot was taken and dried under nitrogen, then dissolved in 2 mL of dichloromethane and 2 mL of deionized water. After thorough mixing by vortexing, phases were separated by centrifugation (5,000 g, 5 min) and the organic phase was recovered, dried under nitrogen, re-suspended in 1 mL isopropanol and filtered through a 0.45 μm PTFE filter previous to HPLC injection.

Acetogenin Chromatographic Determinations

Method I. Extracts were separated with a C18 column (Zorbax Extend-C18, 3x100mm, 3.5μm; Agilent, CA, USA) using a HPLC-VWD (Series 1100; HP, CA, USA) system and a gradient elution program that included water (A) and methanol (B) as mobile phases, and a column temperature set to 25ºC. The gradient, based on previous work,[18] started with a linear gradient from 50% to 60% B during the first seven minutes, with a flow rate of 0.3mL/min, followed by an increase to 80% B from minutes 7 to 10, and then underwent linear increases in both flow rate, from 0.3 to 0.5 mL/min, and in methanol concentration, from 80 to 85% B, from minutes 10 to 30; finally, from minutes 30 to 33, methanol concentration was increased to 100%, and maintained isocratic until minute 40, where initial conditions were reset, and column was equilibrated for 10 minutes prior to next injection.

Method II. Samples were also re-injected using a second separation method because Persin and Persenone B (peaks **(7)** and **(8)**, respectively, and shown in Figure 1) were unresolved under the above-mentioned chromatographic conditions; and Persin has a very low absorbance at 220nm. Method adjustments were made in order to quantify both compounds, and study the effects of the co-elution on

the interpretation of the data from *Method I*. Persenone B was quantified at 220nm, and an extra chromatographic run was made at 208nm (a wavelength in which Persin absorbs the most) with a different gradient program to obtain better peak separation. The modified method consisted of loading samples (5µL) in a longer, polar end-capped C18 column (Synergy Hydro RP, 4.6x250mm, 4µm; Phenomenex, CA, USA) kept at 35ºC. Peaks were separated with a water:methanol gradient, at 1mL/min, starting with 80% methanol, with linear increases reaching 95% and 100% at 15 and 25 minutes, respectively, and returning linearly to 80% by minute 30, with 10 minutes to equilibrate between injections. Chromatogram acquisition was performed in an HPLC-DAD (1525 system; Waters Co., MA, USA) measuring absorbance spectra from 190 to 400nm. Chromatographic profiles were therefore obtained by measuring absorbance at 220 nm for the other acetogenins and 208 nm for Persin.

Identification and Quantification Methods. Chemical identities, under both methods, were assigned by comparing retention times to those of NMR-confirmed, purified peaks by Rodríguez-Sánchez *et al.*[11] Analytical standards (purity >97%) were obtained by partitioning acetone extracts in heptane:methanol and enriched by Centrifugal Partition Chromatography (CPC); fractions were then further purified twice by HPLC in preparative scale and individual compounds were characterized with the aid of 1H and 13C NMR.[11] Calibration curves were generated for every purified compound based on weight, except for Persenone A, for which an extinction coefficient was available. Only peak **(3)**, an Unknown Putative Acetogenin (UPA), was quantified in Persenone A equivalents. Identity confirmation of peaks, for which NMR data was not available, was performed in an HPLC-ESI-TOF system (LC/MSD, Agilent series 1100) using *Method I* settings with acidified (0.1% formic acid) mobile phases. ESI drying gas (nitrogen) was set to 13 L/h, at 250ºC, with a nebulizer pressure of 35 psig; capillary voltage was set to 3 kV and the optical parameter were set to 250, 225, and 60 V for the octopole radio frequency voltage (Oct RFV), fragmentor and skimmer voltages, respectively. Mass spectra was

acquired in positive mode and saved in centroid mode, with an m/z range from 80-1,500 Th, and reading at 1.02 cycles per second.

Data Analysis

All statistical and data analysis were made using the R programming language.[19] Analysis of Variance (ANOVA), t-tests, and Principal Component Analysis (PCA) were performed using the *stats* library[19] and grouping by Tukey's Honestly Significant Difference (HSD) test was done with the aid of the *agricolae* package in R.[20] Effects were considered significant if the p-value was less than 0.05 for ANOVA and less than 0.025 for one-tailed t-tests; an alpha (α) of 0.05 was considered for Tukey's HSD. Canonical Correlation Analysis (CCA) was achieved by using the CCA library in R[21]. Unless otherwise stated, concentrations are shown as mean with standard deviation.

Descriptive Linear Discriminant Analysis (dLDA), used for canonical weight analysis was performed on PCA-normed values by means of the *ade4* package in R[22], in which total variance was constrained to a value of 1 to better explain variation of the existing dataset. On the other hand, predictive Linear Discriminant Analysis (pLDA) equations, used for assignation of putative genotypes, were obtained by processing raw individual concentrations with the *DiscriMiner* library,[23] whose algorithms limit the within variance to 1, and therefore external data, as that generated from new extractions, can be properly evaluated in a dataset-independent manner. These classification algorithms were performed on concentrations of seven acetogenins of each tissue, expressed as a column for each individual fruit. Only cultivars with known genetic backgrounds (Guatemalan, Hybrids and Mexicans; with 4, 5 and 6 members respectively) were kept, resulting in a matrix of 45 rows (15 varieties, by triplicate) and 21 columns (7 acetogenin concentrations in peel, pulp, and seed). Both dLDA and pLDA were performed on individual fruits, and final variety assignations of cultivars were decided by prediction of all (unanimity) or two out of three individuals

(majority) belonging to the same group; if no majority is found, the cultivar would be assigned as conflicting assignation.

Results and Discussion

The collection from the CICTAMEX orchard yielded highly diverse fruit samples with peel colors ranging from green, green and red, to black peel. Fruit size ranged from small (8 x 5 cm, length and diameter) to big (15 x 10 cm) with many different shapes (oval, rounded, pear, etc.) as it can be seen in Supplementary Figure 1. Samples were obtained from cultivars belonging to at least two different varieties: Mexican (*P. americana* var. *drymifolia*) and Guatemalan (*P. americana* var. *guatemalensis*) as well as hybrids and uncharacterized lines (marked as 'Unknown' in Supplementary Table 1). The sample set used herein was therefore selected to reflect the existing diversity of avocado fruits.

Avocados have been classified by fruit morphology since pre-Hispanic times, as documented by friar Bernardino de Sahagún ca. 1590, into three main cultivars: *ahuacatl*, with small, black fruits; *quilahuacatl*, with green, savory fruits; and *tlacazolahuacatl*, with big fruits with large seeds.[24] In modern times, varieties were renamed Mexican, Guatemalan, and West-Indian, respectively; however, the botanical classification has barely changed over the centuries, with the same characteristics used to differentiate Mexican fruits (thin skin, large and detached seeds, and nutty flavor) from Guatemalan (thick or woody skin, small and attached seeds, savory pulp) and West-Indian (smooth skin, big fruits, with a pulp not as palatable as the previous two); with the latter used mainly for rootstock, and not consumption.[25] Taking heed of this, the avocado fruits sampled seem to agree well with the morphological classification (Supplementary Figure 1 and Supplementary Table 1). The samples of unknown genotype, however, are more difficult to classify by morphology, with the exception of fruits from Los Catorce (L14Ch and L14NE), which would fall in the Mexican variety given their edible, black skin and detached seed. This is also in part due to the difficulty of classifying hybrids, which has

led to long held misclassifications, such as 'Hass' avocado, which was classified as a pure Guatemalan until recently, when molecular tools classified it as a balanced hybrid (M x G - 42%, 58%).[26]

Chromatographic Profiles and Chemical Identification of Acetogenins

After extraction and purification, all tissues from almost every sample analyzed accumulated at various degrees a total of eight compounds that were separated in seven chromatographic peaks with analytical *Method I* (Figure 1A). Six of these compounds were identified as previously characterized acetogenins, since they matched the retention times, MS adducts and fragmentation patterns of NMR confirmed standards and published results (Figure 1B, Supplementary Table 2, peaks **(2)** and **(4)** through **(8)**). [11, 12, 18, 27] Apart from these previously characterized acetogenins, two other major peaks (**(1)** and **(3)**) were also identified as putative acetogenins. Identity confirmation as acetogenins was based on data from mass spectral analyses as described in methodology section. Analyses indicated that the heaviest calculated m/z values in positive mode were of 347.2418 and 349.2615 for peak **(1)** and **(3)**, respectively (Supplementary Figure 2 and Supplementary Table 2). These masses corresponded to sodium adducts that resulted in possible molecular weights of 325.2535 and 327.2692 when Na^+ was subtracted. In addition, compound fragmentation patterns matched common neutral losses typical of acetogenins,[11] such as acetate (loss of 60 Th) and acetate minus one water molecule (-78 Th). Moreover, the relative intensity of these ions matched that of purified standards of Persenone B (peak **(8)**), which were analyzed along with the samples (Supplementary Figure 2). Putative formulas for peaks **(1)** and **(3)**, were then assigned based on the exact molecular weight and neutral losses, under the hypothesis that they were acetylated acetogenins and thus, contained 4 oxygen atoms. This assumption was made based on the similarities to other acetylated acetogenins (see Supplementary Figure 2 and Supplementary Table 2), particularly on the neutral losses previously mentioned, that accounted for at least one water molecule and an acetate group. Thus,

formulas were assigned tentatively as $C_{19}H_{32}O_4$, and $C_{19}H_{34}O_4$ for peaks **(1)** and **(3)**, respectively. Molecular ion, adduct masses and fragmentation pattern for peak **(1)**, all matched data previously reported for 1-Acetoxy-2,4-dihydroxy-heptadec-12-en-16-yne,[27]; thus peak **(1)** was assigned the same putative identity and labeled as AcO-avocadenyne in Figure 1B. On the other hand, peak **(3)** was labeled in Figure 1 as Unknown Putative Acetogenin (UPA) as it matched mass spectral characteristics of various reported acetogenins (Supplementary Table 2, Figure 1).[27]

Acetogenin Profiles in Avocado Fruits from 22 Different Cultivars

Acetogenin concentration was quantified in the three tissues for all samples obtained to evaluate variations amongst this large genetic pool. Total acetogenin concentrations (TACs) were higher in pulp and seed (Figure 2A and B, respectively) than in peel (Figure 2C), the former two not being statistically different. Peel concentrations were always the lowest in all cultivars, except for 264 PTB, L14NE and 'Aguilar'. In the case of 'Ag. Negro', acetogenin concentrations were higher in peel than in pulp. As for TAC differences between pulp and seed, only six cultivars were statistically different, with three of them ('Comcar 1', 'Reed' and 'Aries') presenting higher concentrations of acetogenins in seed, and other three (L14NE, 'Pionero' and 'Aguilar') presenting an opposite trend. Interestingly, although acetogenin profiles were similar for all cultivars, concentrations of some individual compounds were different between pulp and seed but without an obvious trend, with the exception of Persin, which was consistently higher in pulp from all varieties when compared the other two tissues. In a similar manner, but at lower relative concentration to pulp values, the peel consistently had a high Persin contribution. This outcome, that seed tissue has at least the same concentration of acetogenins than pulp in most cultivars, is particularly promising for production purposes, since seeds are waste from the food industry, and thus can be an inexpensive input for extraction. Moreover, since seeds contain an order of magnitude less oil than pulp (1.9 vs. 15.4%, respectively)[28] extraction of acetogenins from this tissue

232 may be translated in less lipophilic contaminants, (such as fatty acids, triacylglicerols, etc.) and a less
233 extensive cleaning process.

234 Mean separation conducted on TACs from seed tissues (Figure 2B) indicated that most cultivars
235 grouped in a large homogeneous group, with a wide range of mean values (1.1-6.3 mg/g FW).
236 Concentrations were not statistically different in that group due to the high variance encountered in
237 seeds, with an average coefficient of variation (COV) of 44%, and COVs as high as 140%. The 'Aries'
238 cultivar, although it belonged to the same large grouping, contained TACs that exceeded average seed
239 concentrations for slightly more than 2-fold, having a TAC of 8.33±0.622 mg/g FW. This contrasts
240 with the group of statistically discernible cultivars, formed by 'Aguilar', 'Reed', 'Comcar', 'Aquijic',
241 'Larrainzar', L14NE and L14Ch, which presented a lower range of TACs (1.24-2.30 mg/g FW).

242 Pulp grouping by TAC (Figure 2A) resulted in a similar situation as seed: the largest statistically
243 homogeneous group ('f'), contained 16 cultivars, spanned almost an order of magnitude (0.49-4.3
244 mg/gFW) with an average of 2.36±1.39 mg/gFW. The remaining 6 cultivars were divided equally in
245 two groups: discernible and non-discernible from the group with the highest concentration ('a').
246 Correspondingly peel (Figure 2C), the tissue with the lowest TAC levels, had a large group of cultivars
247 which showed no statistically significant differences, comprising 19 cultivars, with a range that spans
248 an order of magnitude (0.22 – 2.8 mg/gFW) and a TAC of 1.1±0.84 mg/gFW, contrasting with the rest:
249 'Ag. Negro' (5.5±1.3 mg/gFW), L14NE (4.6±1.1 mg/gFW) and 264PTB (12.5±3.0 mg/gFW). The
250 latter avocado line, corresponding to a non-commercial accession, presented consistently high amounts
251 of acetogenins in all tissues tested.

252 When acetogenin profiles were compared among tissues, Persenone A always resulted as the main
253 acetogenin in peel and pulp, encompassing 46 and 48%, of total acetogenins respectively. However, for
254 three cultivars ('Vargas', 'Aries' and 'Larrainzar'), Persin was the most abundant acetogenin. For the

peel tissue, there was a larger number of cultivars that had Persin as the most abundant acetogenin. In seeds, however, less than half of the cultivars had Persenone A as the main acetogenin. AcO-avocadene was the main acetogenin in seeds from 'Pionero', 'Aries', 'Encinos', 'Hass', 'Fundación 2' and 'Larrainzar' cultivars, and AcO-avocadenyne was the most abundant compound in other six cultivars (L14NE, L14Ch, 'Ag.Negro', 'Almoloya', 'Aquijic', and 'Comcar 1'). Noteworthy, almost all cultivars in which AcO-avocadenyne was the main contributor to the seed acetogenin profiles were either confirmed or presumed Mexican varieties, with the exception of 'Comcar 1', which was labeled as Guatemalan according to data provided at collection.

The quantitation of these eight compounds across 22 avocado cultivars from Mexican and Guatemalan varieties along with their hybrids (Supplementary Table 1) showed that almost two thirds of the examined cultivars (14 out of 22) contained low contents of acetogenins (Figure 2). Although this may seem counterintuitive from an evolutionary perspective (it would be expected for high acetogenin content to be positively selected, due its role in plant defense,)[13, 14, 29, 30] it may be explained by co-selection of the trait during domestication, since acetogenins have been reported to present a bitter, unpleasant flavor.[3, 31] Therefore, it is possible that varieties with a high concentration of acetogenins may have been selected out. In the light of this rationale, it seems coherent that commercially accepted cultivars such as 'Reed' and 'Hass' have low concentration of acetogenins in pulp, while the highest concentration was found in a non-commercial accession (264PTB, Figure 2). However, when considering the few cultivars that were statistically discernible, average contents of TACs span more than one order of magnitude in peel (0.22-12.5 mg/gFW) and in pulp (0.49-9.6 mg/gFW) and less than one in seed (1.1-8.3 mg/gFW, Figure 2). These observations are in accordance with a previous study conducted in avocado leaves with 21 avocado lines, which revealed that Persin concentrations varied

within a similar range (0.4 – 4.5 mg/g FW)[15]. In the present work Persin concentrations ranged from 0.09-0.9, 0.05-1.3 and 0-0.3 mg/g FW for peel, pulp and seed, respectively).

Classification of Acetogenins in Families Based on Their Carbon Number

Data from individual acetogenin concentrations obtained from all cultivars and tissues were plotted in a ternary diagram, common in the food chemistry field. In order to facilitate biological interpretation, acetogenins were grouped by carbon number of their non-acetoxylated backbones in C17 (AcO-avocadenyne **(1)**, AcO-avocadene **(2)** and the UPA **(3)**), C19 (Persediene **(4)**, Persenone C **(5)**, and Persenone B **(8)**) and C21 (Persenone A **(6)** and Persin **(7)**). The ternary diagram was able to separate seed tissue from peel and pulp (Figure 3A). Peel and pulp profiles were indistinguishable, and were enclosed in the lower-right quadrant, with more than 50% C21 and less than 30% C19 acetogenins. In contrast, seed tissue had more than 25% of the acetogenins belonging to the C17 group, and less than 50% to C21 group, while maintaining less than 30% acetogenins with C19. Thus, if a crude avocado extract has more than 25% of its acetogenins belonging to the C17 group, or less than 50% to C21 backbone, it probably originates from seed. There was not a case in which a tissue had more than 35% of the measured acetogenins belonging to the C19 group.

Since the ternary diagram suggested a role of the number of carbons of avocado acetogenins in their accumulation and therefore capacity to classify seed tissue, the present work proposes acetogenins to be grouped in three families C17, C19 and C21 acetogenins, considering carbon numbers of their de-acetoxylated backbones (Figure 1B). From this arrangement, a descriptive nomenclature is here being proposed, using the roots *avocatl* (Nahuatl for avocado) for 17 carbon-, *pahuatl* (Nahuatl for fruit, still in use for some Mexican North-Eastern avocado cultivars) for the 19 carbon-, and *Persea* (from the genus) for the 21 carbon-containing acetogenins. Therefore, acetogenins are to be separated in three families: Avocatins (C17), Pahuatins (C19) and Persenins (C21). This nomenclature is consistent with

all the known lauraceous acetogenins and most of the published trivial names: Persin and Persenone A belong to Persenins, and all reported avocadenols[32] and avocadenes[31] (acetoxylated or not) would be in the Avocatins family, along with the avocatin B (a mixture of AcO-avocadene and AcO-avocadyne) [10]. It also recovers the use of the term Avocatins, employed in the seminal work that named 17-carbon acetogenins.[4] This nomenclature is to be fully developed to include all known lauraceous acetogenins in further reports.

Effects of Tissue Type on Acetogenin Profiles and Concentrations

To explore the vast number of multi-dimensional data points generated, and discover relationships in an unbiased manner, concentrations of acetogenins (analyzed by *Method I*) were processed by a multivariate, targeted metabolomics approach. As a first approximation, a PCA was performed on the data and results are shown in Figure 3B, which indicates that the two main components were able to explain 94.4% of the variance in the data. As it can be seen, by color-coding of data points, seed acetogenin profiles (red dots) again clearly separated from those of pulp and peel (green and black dots, respectively), which were not discerned among them. Loadings, represented as vectors in Figure 3B, showed that AcO-avocadene **(2)** was the main acetogenin responsible for differentiating seeds from the rest of the tissues. On the other hand, Persenone A, orthogonal in direction, was responsible for grouping peel and pulp. Additional accountability for the variance was given by a third component (5.2%), whose main loadings were AcO-avocadenyne and AcO-avocadene, and also differentiated seed data (Supplementary Figure 3). Therefore, acetogenin profiles from seeds resulted to be particularly characteristic for that tissue and were able to explain as much as 99.6% of the variance of the data set by PCA. These results were also in agreement with observations from the ternary diagram (Figure 3A), and indicate that two C17 acetogenins, AcO-avocadenyne and AcO-avocadene, played a pivotal role

for the differentiation of seed from the other tissues. Persenone A (C21), in contrast, contributes more to the acetogenin profile in peel and pulp (47%) than in seed (30%; $p < 0.001$).

Results from the different approximations converge in establishing that pulp and peel contain similar profiles. CCA indicated that acetogenin concentrations in peel were largely explained by pulp (72% of redundancy) with the variation of a set of acetogenins in pulp explaining the behavior of their exact counterparts in peel, therefore hinting at direct transport. However, peel generally accumulated minute amounts of acetogenins when compared to those from pulp and seed (Figure 2). Our observations suggest that peel tissue cannot synthesize acetogenins and that it relies on transport from the pulp for supply, therefore having a very similar profile. These inferences made from the multivariate analysis are supported by previous works, which noted that peel was incapable of synthesizing Persin from isotopically labeled linoleic acid or acetate[16].

If we try to connect these findings with acetogenin production and possible transport among tissues, this would imply that, while some transport may be involved among the tissues, AcO-avocadenyne seems be locally synthesized in seeds. Pulp is known to be able to synthesize acetogenins,[16] however, seed tissue has not been evaluated for this capacity. The fact that this tissue clearly differentiates itself from the other two raises the hypothesis of being capable of synthesis. Regarding acetogenin biosynthetic route, evidence from previous works indicate that linoleic acid (C18:2) is a precursor of Persin (C21:2)[16]. The vast array of acetogenins found here and their backbone differences, hints to differences in the use of precursors for their biosynthesis. For example, both AcO-avocadenyne and AcO-avocadene are calculated to have a 17 carbon backbone, this implies that their biosynthesis possibly has a different precursor than Persenone A, with 21 Carbon atom backbones (Figure 1B). Persin probably have common precursors with Persenone A, as evidenced by the similarity of their accumulation and structures (particularly the two *cis* double bonds, shared also, in the same ω position,

with linoleic acid, Figure 1B). However, it would be unlikely for Avocatins (such as AcO-avocadenyne and AcO-avocadene) to share linoleic acid as a precursor, not only because of the lack of *cis* bonds in reported avocatins,[31, 32] but also because their carbon number is lower, requiring cleavage and oxidation steps that would make their production from linoleic acid improbable. It is more likely that avocatins' precursors are saturated, medium-to-long chain (C<18) fatty acids.[28] Interestingly, it has been reported that seed tissue stores odd-chain fatty acids of the same carbon number that avocatins in a preferential manner (C17:0, 1.7% and C17:1, 0.37%) compared to pulp (0.033 and 0.11%, respectively) along with other odd-chain fatty acids (C19:0 while not detected in pulp, totals 0.61% of seed fatty acids).[28] Taking all this into account, it is tempting to speculate that seed is capable of independent acetogenin production, and that such biosynthesis may be related to fatty acid availability.

Effect of Variety on Acetogenin Profiles and Concentrations

The relative accumulation of particular acetogenins among the cultivars of known origin and the difficulty for classification of some accessions and hybrids, led us to think that possible the profile of this compounds characteristic of avocado could be used for an initial classification. Potential relationships between different genetic backgrounds and acetogenin profiles were thus explored by a descriptive Linear Discriminant Analysis (dLDA), as described in methodology. For this approach, the model was developed using the cultivars of known origin. The genetic background for 14 of the cultivars (Supplementary Table 1) was provided by CICTAMEX [33], 4 Guatemalan, 5 Hybrids, and 6 Mexican.

As shown in Figure 4A, the main Discriminant Score (DS1) was capable of separating Hybrids from Mexican and Guatemalan varieties, which were almost identical in projection, but formed antipodes on the second dimension (DS2). A canonical weight analysis (Figure 4B) was also performed, which provided insights on the acetogenin profiles that contributed to genotypic classification observed in

Figure 4A. Results indicated that the main discriminant factors were pulp acetogenins (green vectors), followed by peel acetogenins (gray vectors) and, as a distant third, seed acetogenins (red vectors). For example, Hybrids were separated from the other cultivars, mainly by their pulp concentrations for UPA and Persenone C, which had almost no contribution on the separation of Mexican from Guatemalan varieties. In contrast, concentrations of Persenone A and AcO-avocadene in pulp had a major effect on separating Mexican from Guatemalan genotypes, while failing to separate them from hybrid varieties.

It is important to note that, while significance of the loadings in canonical analysis was proportional to their weights, the model and its usefulness depends on the contribution of all coefficients. Hence, in order to reduce the number of components needed for a useful predictive model; a linear regression discriminative analysis was applied to the concentrations matrix data, adding columns by one in each cycle, based on the canonical weights of the dLDA (Figure 4B). Afterwards, the matrix with the reduced number of columns was subject to pLDA with a 3-fold cross validation and 5000 iterations bootstrapping, and the percentage of miss-assignments was calculated for each variety (Supplementary Figure 4).

In order to obtain the combination of weights that resulted in the minimum number of total wrong assignations, the sum of all error matrices was plotted, and a global minimum was found, comprised of the sum of local minima in the single plots (Supplementary Figure 4), when using the best 5 weights for DS1 and 2 for DS2. It was important to note that using more weights was detrimental to the model, as use of less, due to the over-fitting characteristic of LDA models. Also, taking advantage of this global minimum, best fitting results can be obtained by using only one third of the available data.

A reduced dLDA model was then generated, in which pulp and peel had the highest number of corresponding coefficients, capable of accurately describing varieties, and seed tissue didn't appear to contribute in such task. Interestingly, for the majority of the acetogenins, Guatemalan genotypes were

characterized as having significantly less amount of acetogenins than the rest, particularly in peel. This phenotype correlates with the differences in texture and thickness found in the peel in fruits from these two origins: Guatemalan varieties have a thicker, lignified skin while Mexican ones have a soft, even edible peels, in which probably the transport from the pulp that we are suggesting could happen with less constrains. Notably, hybrids showed significantly higher accumulation of the selected acetogenins, suggesting this trait could be heterotic. Average TAC of hybrid varieties doubles their counterparts in pulp (4.75±2.81 vs. 2.36±1.91 mg/gFW) and peel (3.41±4.94 vs. 1.59±1.57 mg/gFW), and almost doubles the TAC in seed of non-hybrid varieties (5.32±2.32 vs. 2.99±1.67 mg/gFW). This observation is relevant, as it opens the possibility of increasing acetogenin content by selective breeding, with a particular emphasis on outcrossing.

Predictive Model for Avocado Varieties based on Acetogenin Profiles

Once the regression LDA was performed the resulting model should be capable of predicting an avocado variety based on acetogenin profiles and concentrations. For this, the reduced pLDA model used only columns matching the most important factors found in the reduced dLDA: concentrations of all acetogenins, with the exceptions of AcO-avocadenyne and Persediene in pulp; and Persenones B and C in peel (seed was not included). The resulting equations for classification are shown in Supplementary Table 3, in which each coefficient was multiplied by the concentration of the corresponding acetogenin in a tissue, and the resulting sum was used as the score for each genotype; the genotype with the highest score was taken as the predicted outcome for the individual sample.

As a validation of the model, prediction equations were applied to avocados of known genotype, correctly classifying all cultivars: 13 by unanimity and 2 by majority (Supplementary Table 1.) The results of the cross validation are summarized in Supplementary Table 4 for individual predictions. Given that it takes two, out of three, individuals sharing a genotype assignation, the adjusted probability

of a prediction to be wrong is 10.4%, 4.6% and 6.9% if the resulting classification is Guatemalan, Mexican and hybrid, respectively. Conversely, the chance of a Guatemalan, Mexican or hybrid variety to be misclassified is 2.3%, 10.4% and 11.4%, in that order. This translates in Mexican and hybrid cultivars having a one in ten probability of being misclassified, probably as Guatemalan.

Genotype prediction for the seven avocados of unknown origin, obtained by a pLDA model, resulted in two unanimous assignments and five by majority, with no conflicting assignation. 'Aguilar' and 'Pionero' cultivars were classified as hybrids, an origin that seems correct, since they were a product of an effort to produce new cultivars by CICTAMEX.[34] Likewise, the classification of both cultivars from "Los Catorce" (L14NE and L14Ch) as Mexican genotypes also corresponded to their morphological characteristics. However, classification of 'Fundación 2' and 'Aries' as having a Mexican origin contrasts with the observed morphology. These particular cultivars do not show phenotypical characteristics of Mexican cultivars, probably being hybrids with a chemical phenotype similar to Mexican avocados, explaining the classification of one of their replicates as Guatemalan and Hybrid, respectively. An increase in sample number is recommended to increase accuracy of the assignation.

Similarly, 'Ariete', the only cultivar assigned as Guatemalan, has been reported to be a segregate of 'Colin V-33' line[35] that is a reported hybrid.[36] However, morphological characterization studies show a strong seclusion from different Mexican avocados,[35] which corresponds to the results of the present study. It is important to note that classification of avocado varieties has been challenging, at best. Often, assignation varies depending on the method used for determination, differing between morphological and molecular biology based techniques. As an example, 'Hass' cultivar has been classified by morphology as Guatemalan[25] but microsatellite markers classified it as an hybrid of Guatemalan and Mexican genotypes.[26, 36] This is the first time that avocado varieties are attempted to

be cataloged considering fruit chemotype as classifying trait, establishing a basis for phenotype-level assignation with well-defined features, that can add up to morphological classification. If enriched by incrementing sample size and tissues analyzed, this method has the potential to be useful for high throughput screenings, since it is easily scalable, faster and cheaper than molecular biology techniques, but also for complementing genotyping with information at a different level, improving robustness of assignations.

Conclusions

In this work six different acetogenins were simultaneously analyzed for the first time in a collection of different avocado genotypes; two compounds were also identified tentatively as aliphatic acetogenins in avocado fruit. Interestingly all acetogenins were present in almost all cultivars screened, contrasting with the marked differences in morphology; thus this metabolism has been conserved in this ancient angiosperm. In addition, multivariate analyses and classification techniques used in this work proved useful in grouping and deducing information of the large data set generated. A good example is the division of acetogenins into three families depending on their carbon number based on the PCA results, which were named as Avocatins (C17,) Pahuatins (C19,) and Persenins (C21.) Furthermore, although peel and pulp have similar acetogenin profiles characterized by Persenone A accumulation, seed profiles were differentiated by the accumulation of AcO-avocadene and AcO-avocadenyne, suggesting biosynthesis capacity by the latter tissue. The usefulness of the proposed classification system in families, also offered the possibility of discriminating tissue extracts based on their acetogenin profiles, with the use of a ternary diagram, regardless of the cultivar. The model generated for classification of avocado varieties based on acetogenin profiles can contribute to the assignation of landraces of unknown origin and also help breeders for material selection.

Acknowledgments

We are greatful to Emeritus Professor Gerónimo Cano and Dr. Juan Carlos Reyes Alemán, from CICTAMEX, for the donation of avocado cultivars. We also thank Rafael Morán and Rodrigo Treviño, from H.E.B. Mexico and Frupack for the commercial 'Hass' avocados.

Carlos Rodríguez-López is working under a CONACyT scholarship (holder number CVU: 256801), and his studies are partially funded by a Tecnológico de Monterrey scholarship and the Micronutrients Research Chair Funds (CAT- 198).

Competing interests

This research was financed by FEMSA Nutrigenomics Chair, and some of the potential commercial uses of acetogenins in the food and pharmaceutical industries are protected under patents WO/2012/042404.

References

1. D. Téliz Ortiz and F. J. Marroquín Pimentel, in *El Aguacate y su Manejo Integrado*, eds. D. Téliz Ortiz and A. Mora, Ediciones Mundi-Prensa, México D.F., 2 edn., 2008, ch. 1, pp. 1-28.
2. M. Yasir, S. Das and M. D. Kharya, *Pharmacognosy Reviews*, 2010, **4**, 77-84.
3. Y. Kashman, I. Néeman and A. Lifshitz, *Tetrahedron*, 1969, **25**, 4617-4631.
4. H. Magalhaes Alves, D. T. Coxon, C. P. Falshaw, W. O. Godtfredsen and W. D. Ollis, *Anais da Academia Brasileira de Ciencias*, 1970, **42**, 46-48.
5. F. Domergue, G. L. Helms, D. Prusky and J. Browse, *Phytochemistry*, 2000, **54**, 6.
6. O. K. Kim, A. Murakami, Y. Nakamura, N. Takeda, H. Yoshizumi and H. Ohigashi, *Journal of Agricultural and Food Chemistry*, 2000, **48**, 7.
7. O. K. Kim, A. Murakami, D. Takahashi, Y. Nakamura, K. Torikai, H. W. Kim and H. Ohigashi, *Bioscience, Biotechnology, and Biochemistry*, 2000, **64**, 4.
8. N. H. Oberlies, L. L. Rogers, J. M. Martin and J. L. McLaughlin, *Journal of Natural Products*, 1998, **61**, 781-785.
9. A. J. Butt, C. G. Roberts, A. A. Seawright, P. B. Oelrichs, J. K. MacLeod, T. Y. E. Liaw, M. Kavallaris, T. J. Somers-Edgar, G. M. Lehrbach, C. K. Watts and R. L. Sutherland, *Molecular Cancer Therapeutics*, 2006, **5**, 2300-2309.
10. E. A. Lee, L. Angka, S.-G. Rota, T. Hanlon, A. Mitchell, R. Hurren, X. M. Wang, M. Gronda, E. Boyaci, B. Bojko, M. Minden, S. Sriskanthadevan, A. Datti, J. L. Wrana, A. Edginton, J. Pawliszyn, J. W. Joseph, J. Quadrilatero, A. D. Schimmer and P. A. Spagnuolo, *Cancer Research*, 2015, **75**, 2478-2488.
11. D. Rodríguez Sanchez, A. Pacheco, M. I. García-Cruz, J. A. Gutiérrez-Uribe, J. A. Benavides-Lozano and C. Hernández-Brenes, *Journal of Agricultural and Food Chemistry*, 2013, **61**, 7403-7411.
12. 2012.
13. C. Rodríguez-Saona and J. T. Trumble, *Journal of Economic Enthomology*, 1996, **89**, 1571-1576.
14. D. Prusky, N. T. Keen and I. Eaks, *Physiological Plant Pathology*, 1983, **22**, 189-198.
15. R. M. Carman and P. N. Handley, *Phytochemistry*, 1999, **50**, 3.
16. A. Leikin-Frenkel and D. Prusky, *Phytochemistry*, 1998, **49**, 8.
17. D. Prusky, I. Kobiler and B. Jacoby, *Journal of Phytopathology*, 1988, **123**, 140-146.
18. D. Rodríguez Sanchez, C. Silva-Platas, R. P. Rojo, N. García, L. Cisneros-Zevallos, G. García-Rivas and C. Hernández-Brenes, *Journal of Chromatography B*, 2013, **942-3**, 37-45.
19. R. C. Team, *Journal*, 2013.
20. F. de Mendiburu, Masters, Universidad Nacional de Ingeniería, 2009.
21. I. González and S. Déjean, *Journal*, 2012.
22. S. Dray and A. B. Dufour, *Journal of Statistical Software*, 2007, **22**, 1-20.
23. G. Sanchez, *Journal*, 2013.
24. B. de Sahagún, *Historia general de las cosas de Nueva España* Imprenta del ciudadano Alejandro Valdés, calle de Santo Domingo y esquina de Tacuba, México, 1829.
25. B. Berg and N. Ellstrand, *California Avocado Society*, 1986, **70**, 11.
26. H. Chen, P. L. Morrell, V. E. T. M. Ashworth, M. de la Cruz and M. T. Clegg, *Journal of Heredity*, 2009, **100**, 9.
27. M. d. R. Ramos-Jerz, Dr. rer. nat., Universität Carolo-Wilhelmina, 2007.
28. P. S. Bora, N. Narain, R. V. M. Rocha and M. Q. Paulo, *Grasas y Aceites*, 2001, **52**, 171-174.
29. D. Beno-Moualem, Y. Vinokur and D. Prusky, *Journal of Plant Growth Regulation*, 2001, **20**, 6.
30. R. Ardi, I. Kobiler, B. Jacoby, N. T. Keen and D. Prusky, *Physiological and Molecular Plant Pathology*, 1998, **53**, 17.

31. A. G. Degenhardt and T. Hofmann, *Journal of Agricultural and Food Chemistry*, 2010, **58**, 10.
32. Y.-C. Lu, H.-S. Chang, C.-F. Peng, C.-H. Lin and I.-S. Chen, *Food Chemistry*, 2012, **135**, 2904-2909.
33. L. López-López, A. F. Barrientos Priego and A. D. Ben-Ya'acov, *Revista Chapingo, Serie Horticultura*, 1999, **5**, 19-23.
34. S. Sánchez Colín and E. Campos Rojas, in *Memoria Fundación Salvador Sánchez Colín*, ed. CICTAMEX, CICTAMEX, Coatepec Harinas, México, 1997, ch. Departamento de Fitotecnia, pp. 113-118.
35. E. Campos Rojas, F. L. Hernández Vasquez, M. C. Espíndola Barquera, J. C. Reyes Alemán and A. Barrientos Priego, Carins, Australia, 2011.
36. M. L. Alcaraz and J. I. Hormaza, *Hereditas*, 2007, **144**, 244-253.

Figure Captions

Figure 1. Acetogenin profile and structures in avocado: (A) Typical HPLC chromatogram, each peak number corresponds to an acetogenin; and **(B)** chemical structures of acetogenins present in avocado fruit.[11] Structures for peaks **2**, **4**, **5**, **6**, **7** and **8** were confirmed by NMR, Peak **1** was tentatively assigned as AcO-avocadenyne based on comparisons of mass and fragmentation spectra with the available literature[27] and **3**, an Unknown Putative Acetogenin (UPA)**,** was assigned a molecular formula of $C_{19}H_{34}O_4$ based on its fragmentation pattern. Since UPA may correspond to three reported acetogenins, which differ only in the position of the insaturations, these are shown as dotted bonds.

Figure 2. Acetogenin contents among avocado cultivars. Acetogenin concentrations are expressed in mg per g Fresh Weight (FW), in peel (**C**), seed (**B**) and pulp(**A**). Means are the average of three biological replicates and error bars reflect the standard error of the total acetogenin content. Letters indicate statistical difference by Tukey HSD test ($p<0.05$). Asterisks mark cultivars for which Persin could not be quantified.

Figure 3. Multivariate analysis for unsupervised classification of acetogenin distribution among tissues. (A) Ternary diagram grouped by carbon number (C17, C19, C21), separated by tissue (seed: red squares; pulp: green circles; and peel: gray triangles); each point is the average of three measurements. **(B)** Average of PCA scores (depicted as points) by cultivar plotted by tissue (seed: red; pulp: green; and peel: gray) and loadings (blue arrows) projected on the two main components (n=3). Size of the arrow is proportional to the magnitude of the loadings; vectors are scaled and therefore, the magnitude does not correspond to the axes; letters indicate the corresponding acetogenin, only shown for relevant loadings.

Figure 4. Model reduction and classification of unknown cultivars. (A) Score plot showing the projection of individuals (each point is a measurement) on the main discriminant planes, showing grouping by genotype (Guatemalan, blue; Mexican, pale green; and Hybrids, orange) and ellipses corresponding to the mean (center) and variance (ellipse) for each class. Density plots are shown along the margins. **(B)** Canonical weights, shown as vectors, corresponding to coefficients of the linear discriminant functions, colored by tissue (seed: red; pulp: green; and peel: gray). Size of the vector indicates normalized magnitude, and labels are located at the mean of each specified genotype; numbers correspond to each acetogenin as shown in Fig. 1. **(C)** The resulting classification using the reduced model, where the center is the mean, ellipses represent the within variance of each group, circles are the known samples used to generate the model and triangles depict the predicted samples of unknown genotype.

Figures

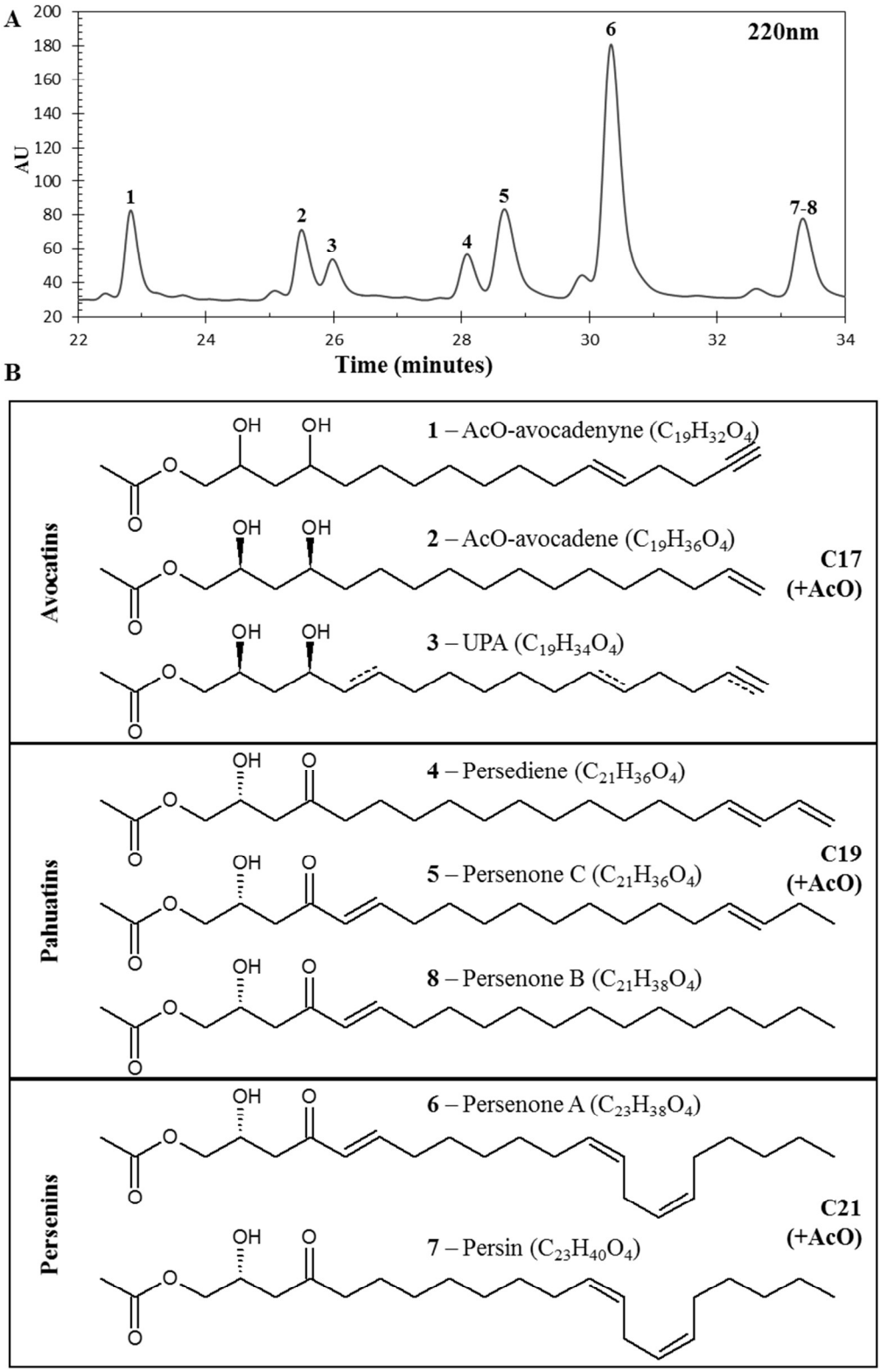

Figure 1.

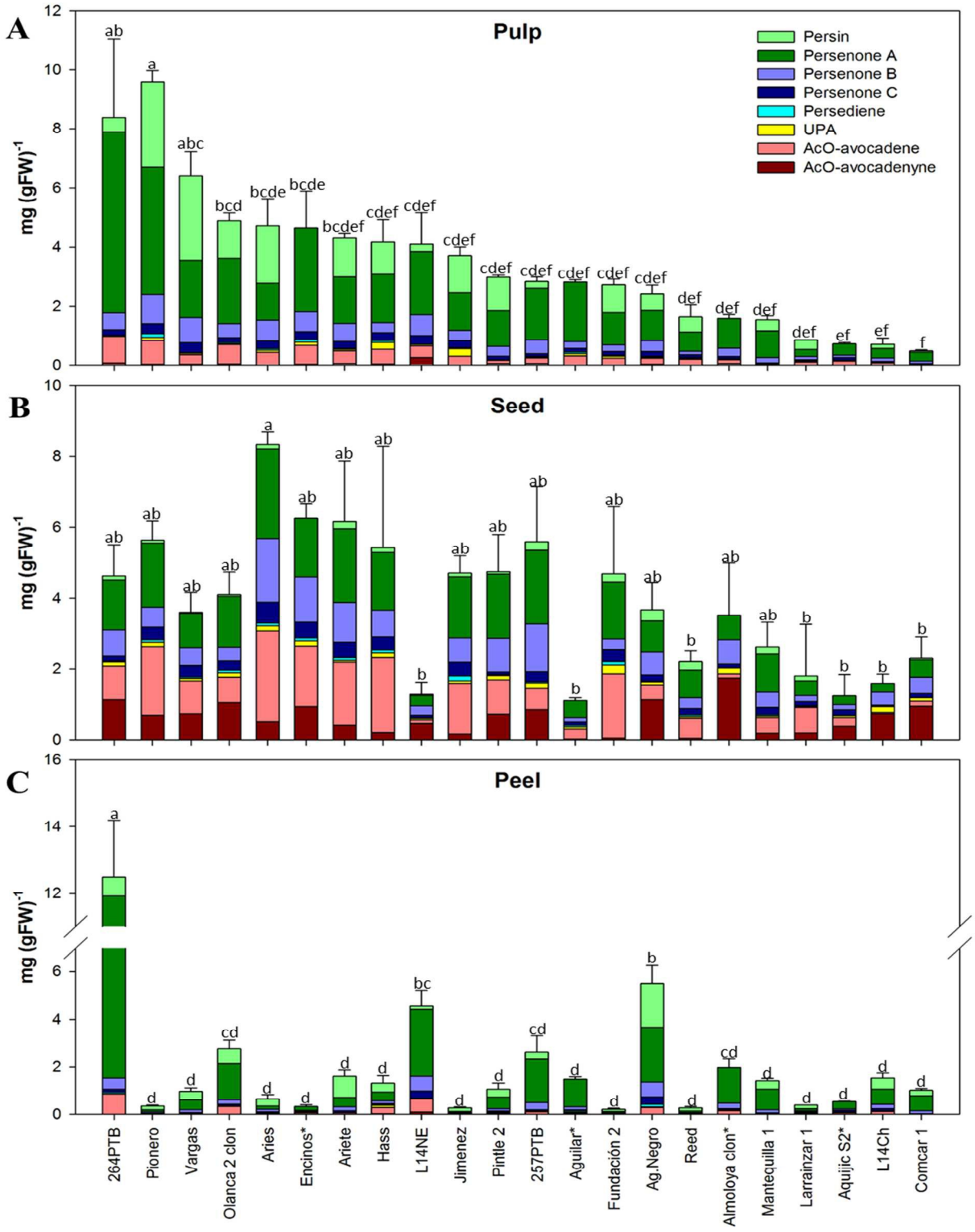

Figure 2.

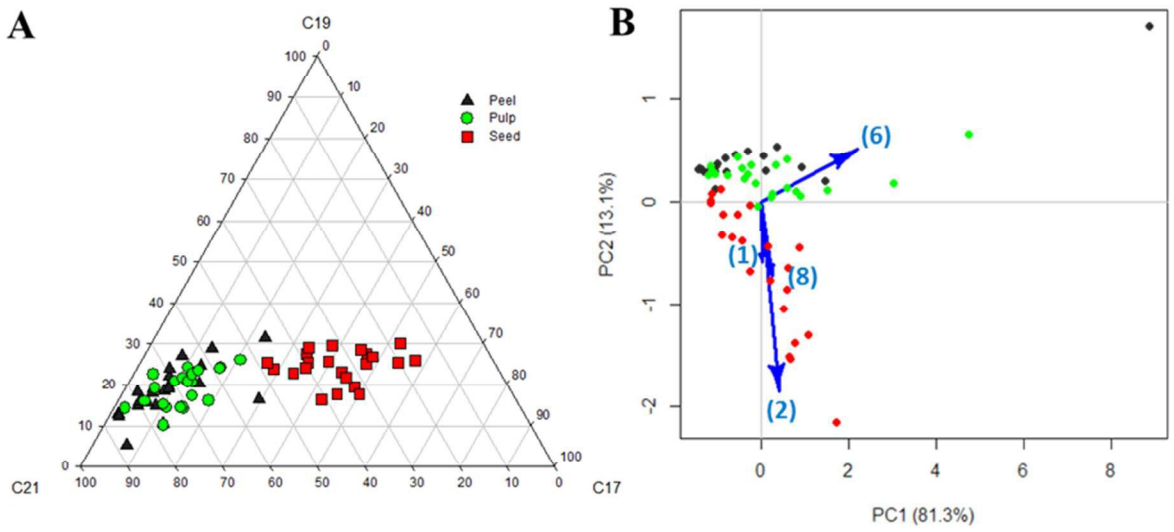

Figure 3.

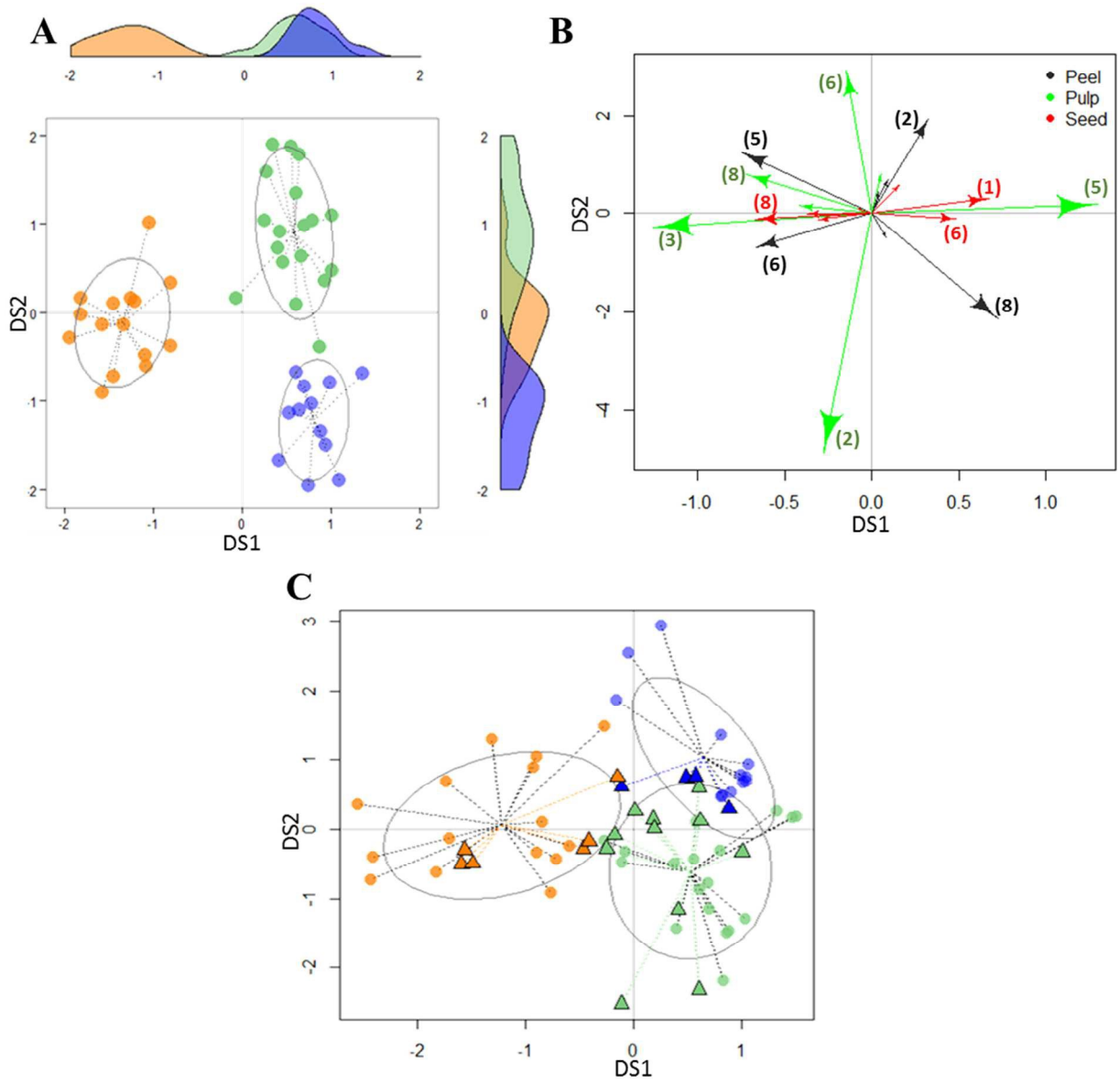

Figure 4.

Laureaceous Acetogenins