JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

Phytochemical Diversity of Cranberry (*Vaccinium macrocarpon* Aiton) Cultivars by Anthocyanin Determination and Metabolomic Profiling with Chemometric Analysis

Paula N. Brown, Susan J. Murch, and Paul Shipley*

Chemistry, University of British Columbia, Okanagan Campus, 3333 University Way, Kelowna, British Columbia V1V 1V7, Canada

ABSTRACT: Originally native to the eastern United States, American cranberry (*Vaccinium macrocarpon* Aiton, family Ericaceae) cultivation of native and hybrid varieties has spread across North America. Herein is reported the phytochemical diversity of five cranberry cultivars (Stevens, Ben Lear, Bergman, Pilgrim, and GH1) collected in the Greater Vancouver Regional District, by anthocyanin content and UPLC-TOF-MS metabolomic profiling. The anthocyanin content for biological replicates (n = 5) was determined as 7.98 ± 5.83, Ben Lear; 7.02 ± 1.75, Bergman; 6.05 ± 2.51, GH1; 3.28 ± 1.88, Pilgrim; and 2.81 ± 0.81, Stevens. Using subtractive metabonomic algorithms 6481 compounds were found conserved across all varietals, with 136 (Ben Lear), 84 (Bergman), 91 (GH1), 128 (Pilgrim), and 165 (Stevens) unique compounds observed. Principal component analysis (PCA) did not differentiate varieties, whereas partial least-squares discriminate analysis (PLS-DA) exhibited clustering patterns. Univariate statistical approaches were applied to the data set, establishing significance of values and assessing quality of the models. Metabolomic profiling with chemometric analysis proved to be useful for characterizing metabonomic changes across cranberry varieties.

KEYWORDS: *cranberries, Vaccinium macrocarpon Aiton, high-performance liquid chromatography, anthocyanins, UPLC-TOF-MS, metabolomics, chemometrics*

■ INTRODUCTION

The American cranberry, *Vaccinium macrocarpon* Aiton (Ericacea), originally native to only eastern North America, has been cultivated in low-lying bogs across North American and parts of Europe,[1,2] with the largest cultivation centers located in Wisconsin, Massachusetts, and New Jersey in the United States and in British Columbia, Canada.[3,4] The American cranberry can be considered a relatively young agricultural crop species, having been domesticated only within the past 160 years.[5] Active breeding and selection of cranberry cultivars are also relatively recent, and unlike many other staple crop species, the cultivated cranberries are little evolved from their wild relatives.[5] It was not until the 1990s that much of the acreage planted with native selections was replaced with first-generation hybrids; however, even today many fields still contain native cultivars.[5]

The first large-scale cranberry breeding program was initiated by the U.S. Department of Agriculture (USDA) in 1929.[5−7] The objective of the program was to develop cranberry cultivars resistant to false blossom disease, a serious disease caused by a phytoplasma carried by the blunt-nosed leafhopper.[5−7] Other selection criteria for the breeding program included yield, fruit rot resistance, keeping quality, fruit appearance, coloring, harvest date, and fruit size.[7] Evidence from other domesticated crops has shown that breeding can significantly affect secondary metabolite levels and diversity within plants.[8−10] These effects are perhaps most prominent in the examination of secondary metabolites that are associated with chemical defense.[9,10] It has been reported that cranberry breeding has had consequences on antiherbivore defense, as evidenced by an increased susceptibility by two cultivars from a selection and breeding program, as compared to three native varieties, to gypsy moth caterpillars.[11] The same study found that the reduced resistance in the bred varieties correlated to a reduction in the levels of several secondary metabolites in the leaves of the plants.[11] Anthocyanins, thought to have several protective biological roles including protection from ultraviolet radiation and solar exposure, cold and drought resistance, and pathogen defense,[12] are reported to vary between cultivars and time of harvest.[13,14]

The current study was designed to characterize and quantify the phytochemical diversity among five cranberry cultivars commonly grown in British Columbia, Canada: Ben Lear, Stevens, Pilgrim, Bergman, and GH1. Ben Lear is a native landrace selection from Wisconsin, which was domesticated around 1900.[15] It ripens early and has very good productivity but is susceptible to rot and other diseases.[15] The Pilgrim, Stevens, and Bergman cultivars are all the result of a USDA cranberry breeding program started in 1929.[5,6] The cultivar Stevens is a cross between the native McFarlin and Potter cultivars and has become one of the most widely used cranberry varieties in cultivation.[5,7] Pilgrim, a cross of McFarlin and Prolific, and Bergman, a cross of Early Black and Searles, are also being grown throughout North America.[7,15] GH1 is a highly productive cross of Rezin McFarlin and Searles that was developed by Ed Grygleski and released in 2004.[15] The breeding origins and relationship between the five cultivars collected, along with an estimate of production percentages in British Columbia, are depicted in Figure 1.

Received: August 18, 2011
Revised: November 28, 2011
Accepted: December 7, 2011
Published: December 7, 2011

 © 2011 American Chemical Society

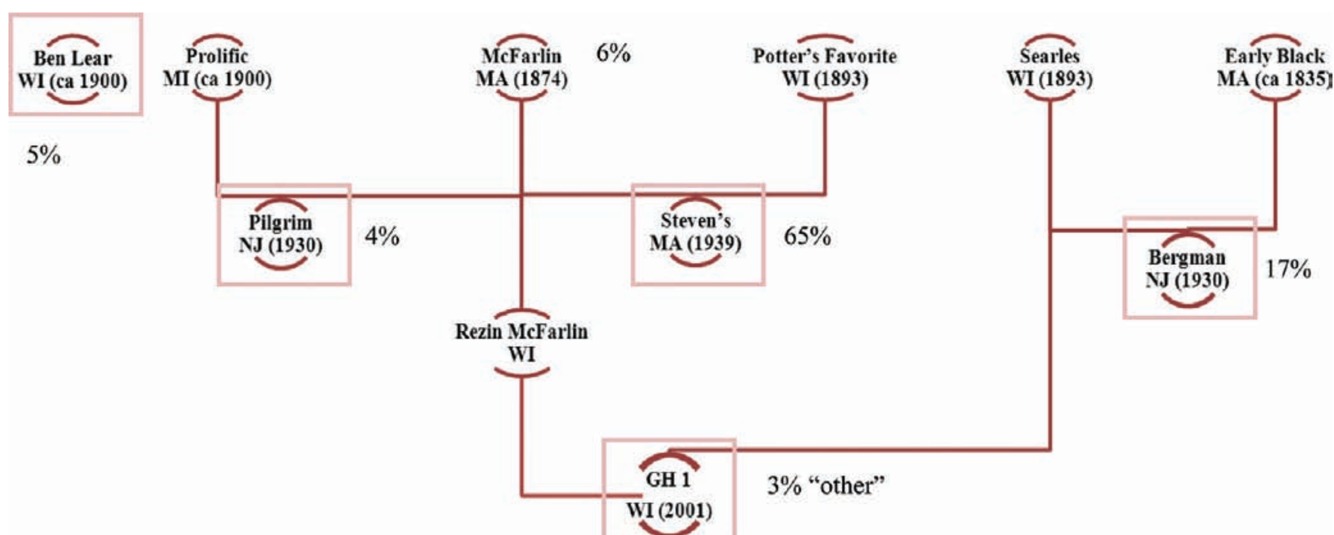

Figure 1. Breeding map showing origins of the five cultivars of *Vaccinium macrocarpon* Aiton being studied in relation to other common cultivars, along with the estimated production percentages of each cultivar in British Columbia.

The specific objectives of the study were to quantify the biological variation of anthocyanin content within and across the cultivars by a previously validated high-performance liquid chromatography with ultraviolet (HPLC-UV) detection method[16,17] adapted to rapid resolution chromatography[17] and to develop a model approach that describes phytochemical diversity and relationships between cultivars through ultra-performance liquid chromatography with time of flight mass spectrometry (UPLC-TOF-MS) metabolomic profiling. Metabolomics, defined as the phytochemistry of metabolites found in an organism being studied within a snapshot of time, can be used to build a picture to represent a system metabolome.[18] Focusing on the entire metabolome, meaning the entire suite of metabolites in a biological system,[19] is a fundamental shift from the characterization of specific classes of plant secondary metabolites. There have been significant advances made in agriculture and food research to distinguish and characterize related materials through the combined use of analytical chemistry techniques to acquire metabolomic profiles with statistical tools to extract relevant information from the data set, referred to as "chemometrics".[8,20−24]

Metabolomics profiles can be compared in a "metabonomic" approach, which is described as the quantitative measurement of time-related multiparametric metabolic responses of multicellular systems to pathophysiological stimuli.[25] A systematic examination of UPLC-TOF-MS metabolomic profiles from commonly produced cultivars of *V. macrocarpon* was undertaken in a process referred to as "subtractive metabonomics".[26,27] This comparison of metabolites from multiple replicates and related biological systems or populations moves beyond characterization at a singular time point to support the identification of relationships across population metabolomes, in this study, cultivars of *V. macrocarpon*. Multivariate data analysis was employed to assist in the evaluation of relationships identified by comparing the metabolomes and visualization of clustering across the cultivars. Different statistical tools were applied to the data set to evaluate data significance and the robustness of the multivairate models developed. This "chemoinformatic" approach introduces a process by which incorrect designation of metabolites as significant and potential misclassifications of relationships can be minimized when considering metabolomic data and offers a new model for interpreting metabolomics data and describing metabonomic changes in the overall phytochemistry of cranberry cultivars.[28,29]

■ MATERIALS AND METHODS

Plant Materials. *V. macrocarpon* Aiton samples were collected from the Lower Mainland of British Columbia from five sites (Figure 2) on October 16, 2010: Ben Lear (N 49 °11.861′, W 123° 02.557′), Pilgrim (N 49° 10.944′, W 123° 00.759′), Stevens (N 49° 11.065′, W 123° 00.523′), Bergman (N 49° 11.760′, W 123° 02.109′), and GH1 (N 49° 11.869, W 123° 02.564′). Cranberry production at all sites was under similar growing conditions (peat) and utilized the same watershed. For all cultivars, the berries were collected immediately prior to the scheduled wet harvesting of the site. Berries were frozen within 4 h of collection and stored at −20 °C until use in chemical and metabolomic analyses.

Chemicals and Calibration Standards. The chemical reagents (hydrochloric acid, glacial acetic acid, phosphoric acid, formic acid) used in this study were of analytical grade and obtained from Sigma-Aldrich Canada (Oakville, ON, Canada). Chromatography and extraction solvents (acetonitrile, methanol, water) were of HPLC grade or equivalent and obtained from VWR International (Edmonton, AB, Canada). Individual anthocyanin chemical calibration standards (250−500 μg/mL) were prepared in 2% concentrated HCl in methanol (v/v) by Cerilliant Corp. (Round Rock, TX) in Snap-N-Shoot format. Mixed chemical calibration standards were prepared for $n = 5$ levels by serial dilution with 2% concentrated HCl in methanol (v/v). The individual Snap-N-Shoot chemical calibration standards were stored at −20 °C when not in use, and the purity was determined chromatographically immediately prior to use.

Sample Preparation. The experiment was designed to compare two different sampling methods, biological and analytical.

Sampling for Biological Variability. A randomized sample selection procedure was followed to subsample five replicate berries from each cultivar collection; each individual berry was weighed, freeze-dried (Modulo Freeze-Dry System; Fisher Scientific, Ottawa, ON, Canada), reweighed to determine percent dry matter, and prepared as individual berries for analysis using a previously published validated analytical method.[16]

*Sampling for Analytical Variability.*Twenty individual cranberries were selected at random from the field collections, combined into a bulk sample, weighed, freeze-dried, reweighed to determine percent dry matter, ground to a <60-mesh powder, and subsampled in five replicates. All cranberry samples were weighed (0.250 ± 0.025 g) into

Figure 2. Collection area (left: red pin) in Lower Mainland, British Columbia, and specific sites of collection for the five *Vaccinium macrocarpon* Aiton cultivars used in this study.

50 mL conical tubes, 20 mL of HPLC grade MeOH−concentrated HCl (98:2, v/v) was added, and samples were mixed with a vortex mixer (Thermolyne Maxi Mix 1; Fisher Scientific) for 10 s, sonicated (Branson model 3510R-MTH Ultrasonic Cleaner; VWR International, Mississauga, ON, Canada) for 15 min, and shaken on an angle at 180 rpm for 30 min (wrist-action shaker model 57040-82; Burrell Scientific, Pittsburgh, PA). The supernatant was decanted into a 25 mL glass volumetric flask, and samples were brought toa final volume of 25 mL with the extraction solvent. Samples were mixed well by inversion, and approximately 1 mL of each sample solution was filtered (0.45 μm Teflon filters; VWR International; Edmonton, AB, Canada) into an amber vial for HPLC analysis.

Anthocyanin Determination by HPLC-DAD. The content of cyanidin-3-*O*-galactoside, cyanidin-3-*O*-arabinoside, cyanidin-3-*O*-glucoside, peonidin-3-*O*-galactoside, and peonidin-3-*O*-arabinoside was determined in the samples as per a published analytical method.[16,17] In brief, anthocyanins were chromatographically separated using an 1100 series Agilent (Mississauga, ON, Canada) liquid chromatography system at ambient temperature using a reverse phase Zorbax SB-C18 Rapid Resolution HT column (4.6 × 50 mm, 1.8 μm) with a mobile composition of (A) 0.5% water−phosphoric acid (99.5:0.5, v/v) and (B) water−acetonitrile−glacial acetic acid−phosphoric acid (50.0:48.5:1.0:0.5, v/v/v/v). The optimized gradient program is 9−36% B over 8.0 min, 36−60% B over 0.5 min, 60−80% B over 0.5 min, and 80−9% B over 0.1 min, with a hold at 9% B for 1.4 min. The injection volume and flow rate were 5 μL and 2.1 mL/min, respectively. The analytes were detected at 520 nm, and data were collected and analyzed using ChemStation software (rev. B.03.01) from Agilent Technologies (Mississauga, ON, Canada). Quantification of anthocyanins was accomplished through the use of standard curves obtained through the analysis of the mixed standard solutions prepared as per the procedure described above. Single-factor ANOVA was used to compare the quantities of the individual and total anthocyanin levels as well as anthocyanin ratios in the biological and analytical replicates between the cultivars. Student *t* tests were used to compare the anthocyanin levels and ratios between the biological and analytical replicates for each cultivar.

Metabolomic Profiling by UPLC-TOF-MS. Cranberry metabolomes were analyzed as per a previously established protocol[23] with an Acquity series ultraperformance liquid chromatography system (Waters Inc., Mississauga, ON, Canada) coupled with a Micromass LCT Premier series TOF-MS (Waters Inc.) and controlled with a MassLynx V4.1 data analysis system (Waters). Chromatographic separation was achieved with a Waters BEH Acquity C_{18}, 2.1 × 150 mm, 1.7 μm column, with the following mobile phase conditions: 0.1% aqueous formic acid/acetonitrile (0.0−10.0 min, 95:5−5:95 v/v, 10.0−15.0 min, 5:95 v/v, 15.0−20.0 min, 5:95−95:5 v/v, 20.0−25.0 min, 95:5 v/v). A 25 min run time was used with a flow rate of 0.25 mL/min and a column temperature of 30 °C. The autosampler was at 4 °C with an injection volume of 5 μL. A Waters 1525 HPLC binary solvent manager provided a steady flow of 2 ng/mL leucine enkephalin at 10 μL/min as the internal calibrant for flight tube length in mass spectrometer.

Data Processing and Subtractive Metabonomics. The metabolomic data of the five *V. macrocarpon* cultivars, acquired by UPLC-TOF-MS experiments, were autoscaled using the standard MassLynx data collection software to set the initial time interval and maximum signal intensity. The resultant data were then compiled with blanks summed and subtracted with any resulting negative values reset to zero in Microsoft Excel (Microsoft Corp., Redmond, WA). A series of automated functions to create subtractive data sets were created in Microsoft Excel using sequential algorithms and functions designated "Subtractive Metabonomics".[26,27] The custom macros designed for this process serve to locate standards, align retention times, remove multiply charged ions, etc., and can provided detailed information from any metabolite data set. After blank subtraction, the data were also set up in an ASCII text file, where rows consist of sample identifiers as objects and columns consist of retention time, m/z ratio, and abundance as variables for import into R (R Foundation, GNU) and Solo+MIA (Eigenvector Research Inc., Wenatchee, WA) for further statistical analysis including multivariate modeling.

Significance Analysis. To assess the degree of significance of metabolites/compounds in comparisons of the metabolomics profiles of the five cranberry cultivars, different univariate statistical tools were employed. First, receiver operator characteristic (ROC) curves were generated and the area under the ROC curve (AUC) were computed for each binary comparison among the five cultivars using the "*colAUC*" algorithm in the "*caTools*" package.[30] The ROC curve is a plot of the sensitivity for a binary classification system.[29] The accuracy of this plot is determined by assessing the AUC, whereby an AUC of 1 would indicate 100% sensitivity at a 0% false-positive rate. To generate a "total" AUC value that captures all possible binary comparisons between the cultivars, the mean of all 10 binary AUC values at each m/z value was determined.[31] In this way the AUC can be used to capture the distribution of abundance across the cultivars and between replicates.

The analysis can become more meaningful when results are supplemented with univariate variance analysis such as the nonparametric Kruskal−Wallis one-way analysis of variance.[32] At each m/z value the Kruskal−Wallis p value was calculated for the replicate abundance data (n = 5 per cultivar) using the "*kruskal.test*" algorithm derived from the "*stats*" package available in R. The p values and total AUC of the

metabolomic data are plotted in Excel to enable identification of those variables that exhibited high variances in abundance across cultivars and small deviations within replicates. This process was employed as a tool to assist in the identification of components (variables) in the metabolomics profiles that may be significant in terms of contributing to cultivar differentiation.

As the AUC is a binary comparison, it is unable to capture interactions or relationships that may exist across the metabolites in the metabolomics profiles; the significance of microarrays (SAM) analysis described by Tusher et al.[33] is another approach for identifying metabolites (variables) of significance that is based on the distribution of abundance in the metabolomics profiles of each cultivar. The SAM statistic was applied to identify m/z values of significance when the five cultivars were compared with respect to abundance; the false discovery rate (FDR) was calculated at each threshold selected.[33] The ranked groups of values and associated FDRs were compiled using the "*siggenes*" package done in R.[34] The m/z values identified by the SAM statistic were distinguished in the plot of p value against total AUC of each metabolite.

Multivariate Analysis. Both unsupervised (PCA) and supervised (PLS-DA) algorithms were selected to observe variance exhibited by the metabolomic data and relate categorical information (sample identifiers) to the abundance data. The only preprocessing of the data prior to the application of PCA and PLS-DA algorithms was autoscaling performed in Solo+MIA. Score plots were generated to visualize clustering by varietal and loadings plots for examining the distribution of data values. Within the loading plots for the PLS-DA models, the calculated Kruskal−Wallis p values for abundances at each m/z value were labeled as two groups: values with a calculated p value >0.05 and values with a p value of <0.05. Those data with p values >0.05 were removed from the data set and the remaining data remodeled. The Q-residual value of each m/z value (variable) was taken from the PLS-DA loading plots.[35] Each Q-residual was transformed into %Q-residual by the formula

$$\left[1 - \frac{Q_{res} - Q_{res}^{min}}{Q_{res}^{max}}\right] \times 100 = \%Q_{res}$$

where Q_{res}^{min} is the minimal Q-residual of the entire metabolomic data set and Q_{res}^{max} is the maximal Q-residual of the entire metabolomic data set. The %Q-residual was used to observe the ability of the chosen model to capture the contributed information existing in the data at each m/z value.

RESULTS AND DISCUSSION

Determination of Anthocyanins in Cranberry Varietals. The Ben Lear, Bergman, and GH1 cultivars had significantly higher anthocyanin levels than the Pilgrim and Stevens cultivars (Figure 3). These trends are in agreement with other studies that reported higher anthocyanin levels in Ben Lear than in Stevens.[14,36] As noted by Vorsa et al., this variation in the anthocyanin levels can be partially explained by the differences in the size of the fruit, as a negative correlation has been observed between fruit productivity and anthocyanin content.[14]

Fruit size differs across the cultivars, with Ben Lear and Bergman having the smallest fruit with cup counts of 70−90 and 65−80, respectively, and Pilgrim and Stevens having the largest fruit with cup counts of 46−66 and 50−60, respectively. Our observations are consistent with anthocyanin concentration being higher in the skin, thereby affecting the surface area to volume ratio and the resultant measured values. With other crops it has been observed that cultivars bred for commercial production have reduced levels of certain secondary metabolites and chemical diversity as compared to wild type.[10,11,37] To allow for high productivity and large fruit size, some trade-offs are expected as additional energy and nutrients are devoted to fruit production.[10,11,37] Both Pilgrim and Stevens were selected in part because of their large fruit size and high productivity, however, this selection may have translated into the inhibition of anthocyanin-producing pathways. Thus, careful consideration and monitoring of several variables is needed in the breeding of cultivars to ensure desired properties are retained and/or improved upon. For example, GH1, a relatively new cultivar, exhibits high productivity without a decrease in anthocyanin production. This cultivar had relatively high levels of anthocyanins, yet is of similar size and is more productive than Stevens and Pilgrim. There are efforts being made to cultivate cultivars that are productive, have large fruit, and have high anthocyanin levels.[7,38]

Figure 4 shows the comparison between the biological and analytical replicates. Although at first glance the biological replicates appear to show higher anthocyanin values, Student t tests indicated that only in Bergman was a statistically significant difference ($p < 0.05$) detected. The biological replicates had a large variance associated with them, whereas the analytical replicates showed a very low level of variance. The results herein presented provide an interesting representation of the trade-offs inherent in composite sampling. The main advantage of composite sampling is a reduction in measurement costs. Analysis of $n = 20$ can thus be accomplished in one measurement using a validated method, and the result provides a good estimate of the population mean. The observed variance from the analysis of the five analytical replicates is predominately due to the inherent precision of the analytical method. The trade-off to this reduced testing approach is in a loss of information concerning the inherent variance of the sample population. As shown in the biological replicates, there is a large variance in the anthocyanin levels between individual berries within a given cultivar population. One implication of the lost information can be seen when the cyanidin glycoside to peonidin glycoside ratios are compared (Figure 4). The analytical replicates of Ben Lear, GH1, and Stevens were found to have lower cyanidin glycoside to peonidin glycoside ratios than Bergman and Pilgrim; however, no significant differences in the ratios were detected when the biological replicates were compared. As the variance in the composite sample is dominated by the method precision and not the variability of the berries, a proper comparison of the ratios cannot be accomplished. This is of particular interest as it is often necessary to pool biological replicates to provide enough material for the analytical technique being utilized. This would have, in this case, potentially led to incorrect conclusions as to differences between the cultivars that actually lie within the range of biological variation. Multiple biological replicates are required to provide the necessary variance information and thus allow for a proper comparison of the populations.

Metabolomic Profiling by UPLC-TOF-MS. The metabolite counts and results of subtractive metabonomics are found in Table 1, which makes direct comparisons across the metabolomics fingerprints of the five *V. macrocarpon* cultivars. Using subtractive metabonomic algorithms described previously,[26,27] 6481 compounds were found conserved across all varietals, with 136 (Ben Lear), 84 (Bergman), 91 (GH1), 128 (Pilgrim), and 165 (Stevens) unique compounds observed. About 55−57% of the phytochemistry described in the metabolomic fingerprints was common across all cultivars, with the exception of the metabolome of Stevens, which shared 65% of the observed phytochemistry with all other cultivars.

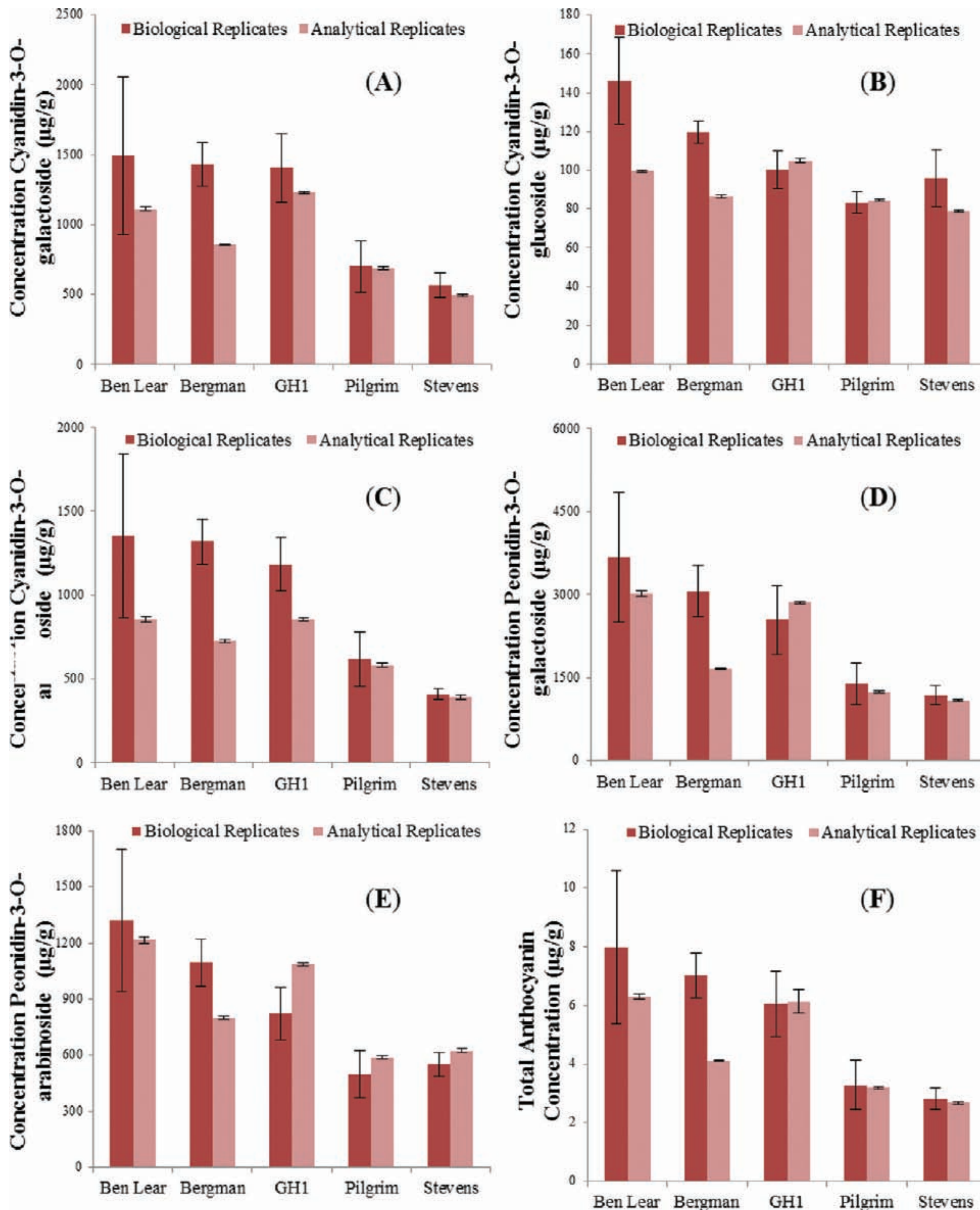

Figure 3. Average anthocyanin content in each cultivar for biological replicates (n = 5) and extractions from a composite sample of cranberries (n = 5) with error bars representing the standard error of the mean for (A) cyanidin-3-*O*-galactoside, (B) cyanidin-3-*O*-glucoside, (C) cyanidin-3-*O*-arabinoside, (D) peonidin-3-*O*-galactoside, and (E) peonidin-3-*O*-arabinosde and (F) the total anthocyanin content in mg/g of dry weight material (DW).

From the binary comparisons of the metabolomics profiles, we observe Stevens and Bergman and Stevens and Ben Lear have the fewest metabolites in common with 83 and 84, respectively, whereas Stevens and GH1 share the most in common with 214.

Chemometric Analysis of Metabolomic Data. In the PCA score plot of the first and second principal components, significant differentiation among the cranberry cultivars from metabolomic profiles was not observed, despite the differences noted in the metabolite counts for each cultivar (data not shown). To observe how the metabolomic profiles of the cultivars may be differentiated, PLS-DA was selected as the next modeling tool (Figure 5A,B). In the score plot modeling linear

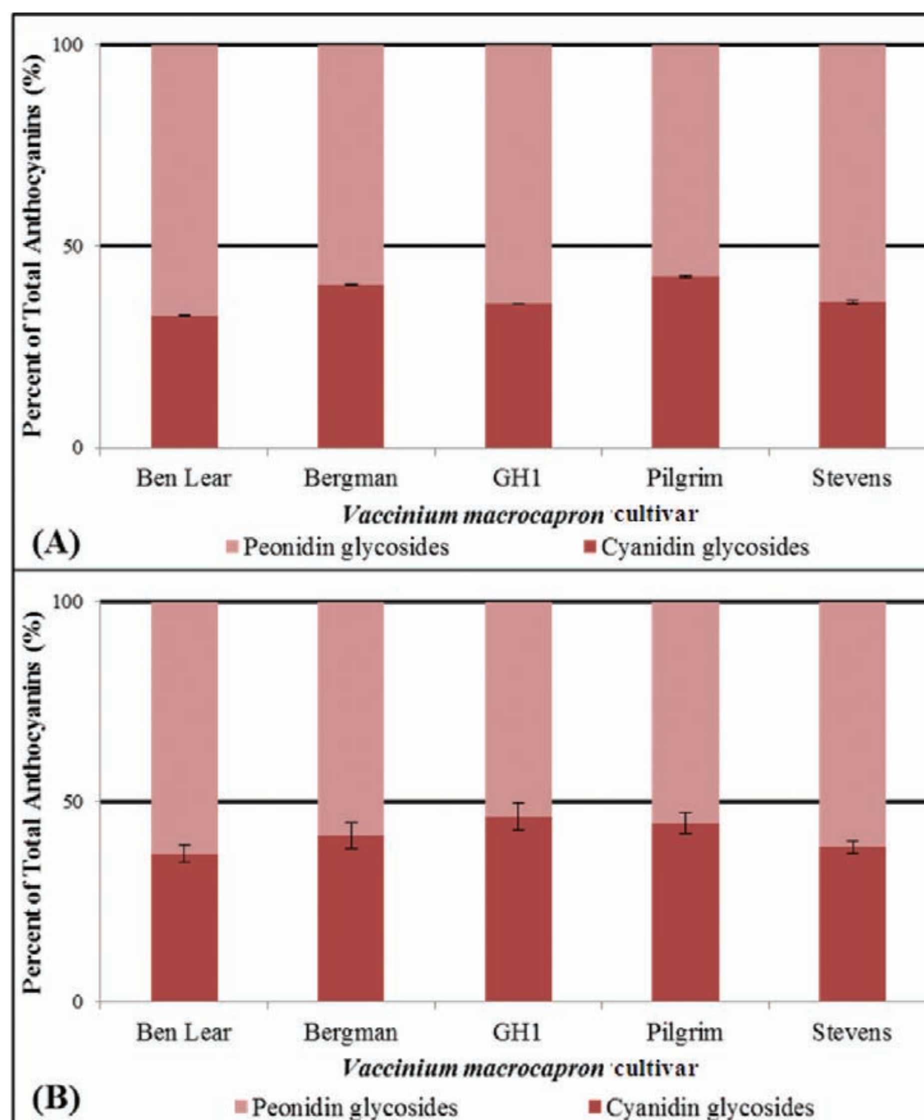

Figure 4. Comparison of the glycosylated peonidins and glycosylated cyanidins in five *Vaccinium macrocarpon* cultivars as a percent of the total anthocyanins in each *Vaccinium macrocarpon* Aiton. cultivar: (A) analytical replicates (n = 5); (B) biological replicates (n = 5). Graphs show averages with error bars representing the standard error of the means.

Table 1. Summary of LC-MS-TOF Metabolite Counts and Differences Metabolomic Profiles of Five Cultivars of *Vaccinium macrocarpon*

	*V. macrocarpon*cultivar				
	Ben Lear	Bergman	GH1	Pilgrim	Stevens
total no. of compounds observed	11544	11395	11322	11736	10038
av no. of compounds per cultivar	5971 ± 195	5944 ± 131	5848 ± 147	6330 ± 222	4477 ± 945
compounds in all replicates	1717	1828	1616	2108	252
compounds in at least 50% of replicates	5254	5262	5225	5674	3765
unique to each cultivar	136	84	91	128	165
common to all cultivars	6481				
common between cultivars					
Ben Lear		152	100	171	85
Bergman	152		96	140	83
GH1	100	96		132	214
Pilgrim	171	140	132		118
Stevens	85	83	214	118	

variates one (LV1) and two (LV2) differentiation of Stevens and GH1 from the other cultivars is observed with the 95% confidence boundaries around Bergman, Ben Lear, and Pilgrim overlapping (Figure 5A). The linked loadings plot (Figure 5B)

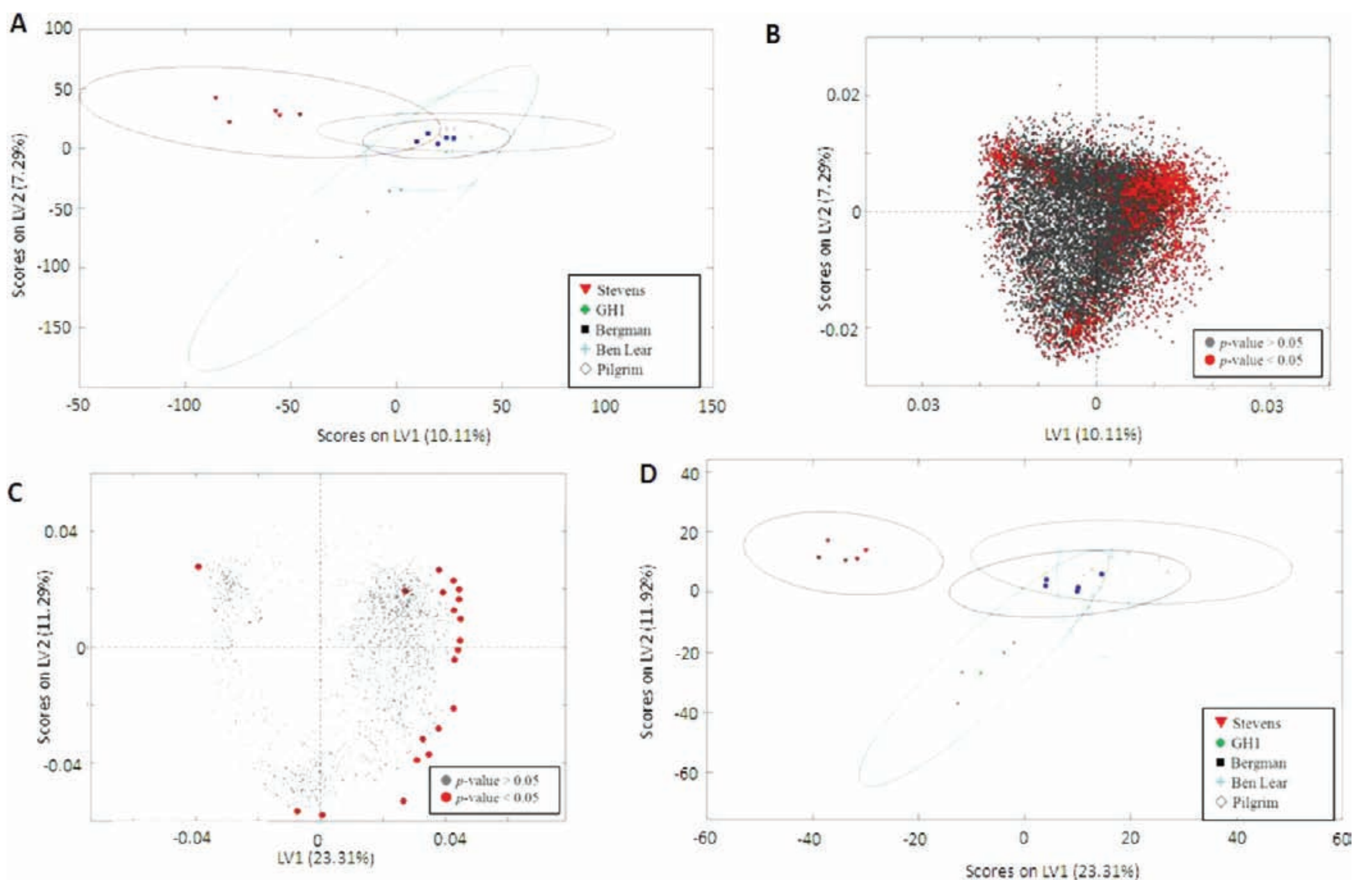

Figure 5. Multivariate analysis of the UPLC-TOF-MS metabolomic profiles of the cranberry varietals: (A) PLS-DA score plot for LV1/LV2; (B) corresponding loadings plot for the PLS-DA model (data with calculated Kruskal−Wallis p value <0.05 are highlighted in red); (C) loadings plot remodeled with all values having a p value >0.05 removed and values identified as significant by SAM statistic highlighted in red; (D) corresponding remodeled PLS-DA score plot for LV1/LV2.

indicates minor clustering of the data by cultivar: Stevens (upper right corner), GH1 (bottom center), and the intense cluster of metabolites where Bergman, Ben Lear, and Pilgrim are located (upper left corner). This clustering is also evident in the values with a Kruskal−Wallis p value of <0.05 (highlighted in red), which indicates the data contributing to the clustering pattern are statistically significant. To further evaluate the relationships between the cultivars and robustness of the data in LV1/LV2, metabolites with p value of >0.05 were peeled off the PLS-DA model and the score and loadings plots remodeled (Figure 5C,D). For the metabolite total of 15064 values, 1749 had a calculated $p < 0.05$, resulting in a loss of 88.4% of the data.

With a limited number of samples ($n = 5$) evaluated per *Vaccinium* sp., the suggested way of validating or testing the performance of the multivariate model by randomly assigning 25% of the samples as "unknown" to test against a model built of the remaining data[39] is not possible, so another approach must be taken. Just as an analytical method can be orthogonally validated by employing an independent method, the results of the PLS-DA model can be compared to a univariate statistical analysis, the significance analysis of microarray (SAM) statistic. The values identified as significant by the SAM statistic are identified (in red) in the LV1/LV2 loadings plot (Figure 5C) to observe whether they fall within the clusters that define the positioning of the five cultivars in the remodeled LV1/LV2 score plot (Figure 5D). With a FDR of 1.78%, the 20 values identified by the SAM statistic are unlikely to be false positives (Table 2).

Table 2. Summary of LC-MS-TOF Metabolomic m/z Values Identified by the SAM Statistic in Order of Ranked Significance, $d(i)$, at a False-Discovery Rate of 1.78% as well as the Q-Residual Values of the Compounds in the PLS-DA Model for Up to LV2 and Up to LV4

		Q-residual (%)	
m/z value	$d(i)$ value	up to LV2	up to LV4
296.385	42.9	98.7	97.7
285.291	26.2	42.4	98.4
305.296	25.8	97.1	100.0
294.354	23.9	95.0	96.5
792.901	23.6	93.8	93.9
482.553	23.4	97.9	97.4
584.593	22.7	83.7	96.3
631.546	22.1	99.0	97.9
398.494	21.9	75.2	80.9
280.337	21.9	80.3	89.6
546.612	20.1	98.5	95.9
585.555	19.4	90.1	89.7
986.927	19.0	85.3	83.5
719.697	18.5	83.7	86.8
398.988	18.4	80.1	87.1
233.316	17.0	91.0	88.8
245.276	16.7	96.6	95.3
276.317	16.6	94.6	93.4
736.798	15.8	89.7	88.6
553.652	15.6	87.9	87.3

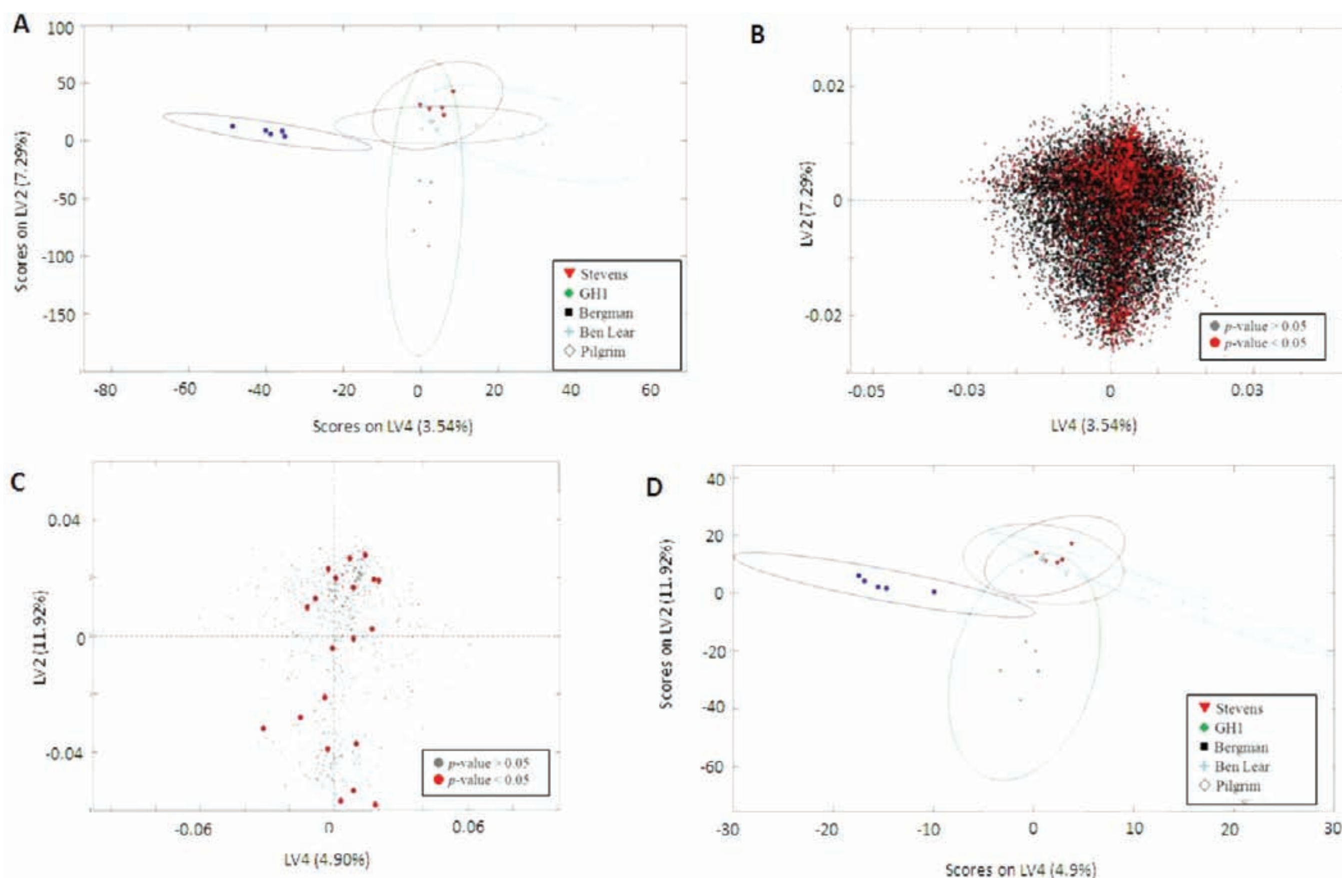

Figure 6. Multivariate analysis of the UPLC-TOF-MS metabolomic profiles of the cranberry varietals: (A) PLS-DA score plot for LV4/LV2; (B) corresponding loadings plot for the PLS-DA model (data with calculated Kruskal−Wallis p value <0.05 are highlighted in red); (C) loadings plot remodeled with all values having a p value >0.05 removed and with values identified as significant by SAM statistic highlighted in red; (D) corresponding remodeled PLS-DA score plot for LV4/LV2.

The SAM values positioned within the clustered areas and at their outermost edges where the contribution of the values to the score plot clustering is greatest (Figure 5C) are supportive of the PLS-DA model. For example, the SAM algorithm identified a m/z value of 276.3168, observed only in the Stevens cultivar, as significant, and this value appears on the edge of the cluster in the loading plot that represents the Stevens cultivar in the PLS-DA model. Similarly, m/z values of compounds observed only in the metabolomics profiles of the GH1 cultivar should be in the lower bottom corner of the loadings plot (Figure 5C), near the LV1 value of zero, and in this region two m/z values, 986.9273 and 398.9882, were identified as significant by the SAM statistic. Although m/z 986.9273 is observed only in GH1, m/z 398.9882 was observed in both GH1 and Ben Lear cultivars. Examining the score plot (Figure 5D), we see that Ben Lear is positioned closest to GH1 and as such the PLS-DA model for LV1/LV2 is again consistent with the SAM statistic.

Often differentiation is not observed in the score plot of the first two linear variates, and other combinations can be considered. Figure 6A is the score plot of LV4 and LV2, which shows a very different clustering pattern from that in Figure 5A. Ben Lear, GH1, and Bergman are somewhat differentiated, and the 95% confidence intervals around Stevens and Pilgrim cultivars are overlapping; however, the linked loadings plot (Figure 6B) shows that values with $p < 0.05$ are not concentrated in the same way but rather are spread in the LV4/LV2 model. When the values with a calculated $p > 0.05$ are removed, the remodeled PLS-DA loadings (Figure 6C) and score (Figure 6D) plots do not exhibit the same degree of clustering. It is expected that LV4/LV2 would not contain as much exhibited variance from the metabolomic profiles as LV1/LV2, but interestingly once the p value <0.05 cutoff is applied to LV1/LV2, there is a 2.0-fold increase in the amount of variance explained (from 17.40 to 35.23%), and for LV4/LV2 the amount of variance explained by the model sees a 1.6-fold change from 10.83 to 16.82%. Looking at the distribution of the values identified as significant by the SAM statistic in the remodeled loadings plot (Figure 6C) shows the 20 values, highlighted in red, cluster in a fashion that loosely reflects the differentiation observed in the score plot (Figure 6D) but are not concentrated at the outermost edges of the LV4/LV2 loadings plot.

On the basis of the combination of orthogonal statistical approaches for evaluating the metabolomics profiles of the five cultivars, we can have more confidence in PLS-DA model of LV1/LV2, which indicates the metabolome of the Stevens cultivar is distinct from Ben Lear, Bergman, GH1, and Pilgrim, although more similar to GH1, which also exhibits a phytochemical signature distinct from Ben Lear, Bergman, and Pilgrim. Although only the LV1/LV2 plots (score and loading) agree with the orthogonal univariate statistics, some merit is seen in the LV4/LV2 plots (Figure 6C,D), where a

Table 3. Observed Average Abundance ($n = 5$), Total AUC, and Ranking by $d(i)$ Value of the m/z Values Identified by the SAM Statistic as Significant at a False-Discovery Rate of 1.78% Compared across the Five *V. macropcarpon* Cultivars

	av abundance for n = 5 replicates						
m/z value	Stevens	GH1	Bergman	Ben Lear	Pilgrim	total AUC	rank, $d(i)$ value
Observed in All Cultivars							
792.901	0.074	0.270	0.360	0.483	0.494	0.920	5th, 23.6
296.385	0.107	1.112	0.627	0.806	0.678	0.904	1st, 42.9
305.296	0.446	0.984	1.381	1.637	1.322	0.896	3rd, 25.8
482.553	0.347	0.482	0.908	0.958	1.121	0.896	6th, 23.4
280.337	0.223	1.205	1.516	1.256	0.950	0.880	10th, 21.9
736.798	1.787	7.160	5.645	6.995	6.331	0.852	19th, 15.8
719.697	0.208	1.235	1.092	1.239	0.892	0.836	14th, 18.5
585.555	0.277	1.494	1.693	1.668	1.472	0.788	12th, 19.4
546.612	0.208	1.995	2.134	2.316	2.273	0.782	11th, 20.1
Observed in Four Cultivars							
584.593		0.038	0.077	0.071	0.123	0.928	7th, 22.7
631.546		0.026	0.061	0.077	0.093	0.912	7th, 22.1
245.276		0.138	0.278	0.264	0.317	0.844	17th, 16.7
233.316		0.383	0.546	0.602	0.605	0.796	16th, 17.0
Observed in Three Cultivars							
398.494			0.052	0.061	0.084	0.884	9th, 21.9
294.354			0.043	0.042	0.058	0.866	4th, 23.9
553.652			0.042	0.060	0.061	0.834	20th, 15.6
Observed in Two Cultivars							
398.988		0.036		0.020		0.844	15th, 18.4
285.291				0.035	0.150	0.784	2nd, 26.2
Observed in Only One Cultivar							
986.927		0.029				0.736	13th, 19.0
276.317	0.115					0.700	18th, 16.6

cluster of the SAM identified metabolites is localized at the outermost edges of a cluster in the loadings plot that reflects where Stevens and Pilgrim are positioned in the score plot. This indicates those metabolite values are truly more significant than other metabolites identified by the SAM statistic with respect to the LV4/LV2 PLS-DA model. Although generally it is assumed with the inclusion of more latent variables in a PLS-DA model the more variance the model exhibits per metabolite value, the comparison of the Q-residuals for up to LV2 and LV4 shows nearly 50% of the SAM identified values (9 of 20 total) reflect a decreasing %Q-residual with increasing number of latent variables (Table 2), thereby providing further confidence in the LV2/LV4 PLS-DA model.

The majority of the SAM identified metabolite values are present in at least three of the cultivars (Table 3). It is interesting to note that where the values are observed in only four, three, or two cultivars, they were not observed in the Stevens cultivar. In fact, outside those values found in all five cultivar, only one m/z value of significance was identified by SAM, m/z 276.317, and observed in Stevens (Table 3). Further investigations on those values not conserved in the Stevens cultivar could lead to a better understanding of the differences and relationships between the native and hybrid cultivars, such as Stevens and GH1.

Applying univariate statistics as a quality assessment of the PLS-DA models developed from the metabolomics data indicates that the LV1/LV2 model is a better fit than LV2/LV4, where the cultivars are further differentiated. Having the majority of the %Q-residual per metabolite value not increasing with an increased number of latent variables further suggests the PLS-DA models are not entirely reflective of the cranberry cultivar metabolomes. However, the metabolomic data and PLS-DA models should be carefully considered as it is expected that the differences and similarities of cultivars from a single species would be subtle. This is further illustrated in Table 3, where the majority of the SAM identified metabolite values exist in all or the majority of cultivars, indicating that significant values existing only in one cultivar are difficult to find. Strategies exist to better improve the "detection" of important metabolites, although for a cultivar study it may prove more useful to increase the sample size to better illustrate the true picture, decrease the value of the associated FDR with statistical models, and cross-validate the multivariate models developed.

Both the targeted and untargeted analyses found significant phytochemical differences among the cultivars. If the goal of the commercial production is yield of anthocyanins, then mass plantings of high-anthocyanin cultivars such as Ben Lear, Bergman, or GH1 would be recommended. However, other agronomic considerations such as yield and disease resistance may be reflected in the large-scale production of Stevens by commercial farmers. It is interesting that GH1, a modern cultivar which exhibits both high productivity and high anthocyanin content, is reasonably differentiated by both PLS-DA models and has metabolites of significance identified in Table 3 both independent of the other cultivars and in common with Ben Lear, Bergman, and Pilgrim. Overall, the

dx.doi.org/10.1021/jf2033335 | J. Agric.Food Chem. 2012, 60, 261−271

significance analyses identified individual metabolites characteristic of each of the cultivars, and a new approach to quality assessment of multivariate models of metabolomics data has been demonstrated. One of the key difficulties with metabolomics data sets is the need for methods and statistical approaches to ensure the quality of data sets. The combined application of the univariate approaches, statistical Kruskal–Wallis *p* values, area under the ROC curve, and SAM statistic, provides the basis for establishing quality evaluations of multivariate models as a new tool for metabolomics research with broad future applications.

AUTHOR INFORMATION

Corresponding Author

*Phone: (250) 807-8749. Fax: (250) 807-8005. E-mail: Paul.Shipley@ubc.ca.

REFERENCES

(1) Suda, J.; Lysák, M. A. A taxonomic study of the *Vaccinium* Sect. *Oxycoccus* (Hill) W. D. J. Koch (Ericaceae) in the Czech Republic and adjacent territories. *Folio Geobotanica* **2001**, *36*, 303–320.

(2) Vander Kloet, S. P. The taxonomy of *Vaccinium oxycoccus*. *Rhodora* **1983**, *85*, 1–44.

(3) Kuhnlein, H. V.; Turner, N. J. Descriptions and uses of plant foods by indigenous peoples. In *Traditional Plant Foods of Canadian Indigenous Peoples: Nutrition, Botany, And Use*; Gordon and Breach Publishers: Amsterdam, The Netherlands, 1996; pp 173–174.

(4) McCown, B. H.; Zeldin, E. L. *Vaccinium* spp. Cranberry. In *Biotechnology of Fruit and Nut Crops*; Biotechnology in Agriculture Series; Litz, R. E., Ed.; CABI Publishing: Oxfordshire, U.K., 2005; pp 247–261.

(5) Hancock, J. F.; Lyrene, P.; Finn, C. E.; Vorsa, N.; Lobos, G. A. Blueberries and cranberries. In *Temperate Fruit Crop Breeding Germplasm to Genomics*; Hancock, J. F., Ed.; Springer: East Lansing, MI, 2008; pp 115–150.

(6) Dana, M. N. Cranberry cultivar list (*Vaccinium macrocarpon*). *Fruit Varieties J.* **1983**, *37*, 88–95.

(7) Vorsa, N. Breeding the American cranberry. In *Wisconsin Cranberry School 1994 Proceedings: Breeding the American cranberry*; Board of Regents of the University of Wisconsin System: Madison, WI, 1994; pp 1–4.

(8) Fernie, A. R.; Schauer, N. Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet.* **2009**, *25*, 39–48.

(9) Hopkins, R. J.; van Dam, N. M.; van Loon, J. J. A. Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annu. Rev. Entomol.* **2009**, *54*, 57–83.

(10) Wink, M. Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores. *Theor. Appl. Genet.* **1988**, *75*, 225–233.

(11) Rodriguez-Saona, C.; Vorsa, N.; Singh, A. P.; Johnson-Cicalese, J.; Szendrei, Z.; Mescher, M. C.; Frost, C. J. Tracing the history of plant traits under domestication in cranberries: potential consequences on anti-herbivore defences. *J. Exp. Bot.* **2011**, *62*, 2633–2644.

(12) Chalker-Scott, L. Environmental significance of anthocyanins in plant stress responses. *Photochem. Photobiol.* **1999**, *70*, 1–9.

(13) Viskelis, P.; Rubinskienė, M.; Jasutienė, I.; Šarkinas, A.; Daubaras, R.; Česonienė, L. Anthocyanins, antioxidative, and antimicrobial properties of American cranberry (*Vaccinium macrocarpon* Ait.) and their press cakes. *J. Food Sci.* **2009**, *74*, C157–C161.

(14) Vorsa, N.; Polashock, J.; Cunningham, D.; Roderick, R. Genetic inferences and breeding implications from analysis of cranberry germplasm anthocyanin profiles. *J. Am. Soc. Hortic. Sci.* **2003**, *128*, 691–697.

(15) Agriculture Research Service NCGR-Corvallis: *Vaccinium* catalog cranberry cultivars and selections, 2009.

(16) Brown, P. N.; Shipley, P. R. Determination of anthocyanins in cranberry fruit and cranberry fruit products by high-performance liquid chromatography with ultraviolet detection: single-laboratory validation. *J. AOAC Int.* **2011**, *94*, 459–466.

(17) Brown, P. N.; Shipley, P. R.; Murch, S. J. Anthocyanin diversity within a commercial production of cranberries (*Vaccinium macrocarpon* Aiton) in British Columbia. *Phytochem. Anal.* **2011** (in review).

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

(19) Fiehn, O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics* **2001**, *2*, 155–168.

(20) Bertram, H. C.; Weisbjerg, M. R.; Jensen, C. S.; Pedersen, M. G.; Didion, T.; Petersen, B. O.; Duus, J. Ø.; Larsen, M. K.; Nielsen, J. H. Seasonal changes in the metabolic fingerprint of 21 grass and legume cultivars studied by nuclear magnetic resonance-based metabolomics. *J. Agric. Food Chem.* **2010**, *58*, 4336–4341.

(21) Cellini, F.; Chesson, A.; Colquhoun, I.; Constable, A.; Davies, H. V.; Engel, K. H.; Gatehouse, A. M. R.; Kärenlampi, S.; Kok, E. J.; Leguay, J.-J.; Lehesranta, S.; Noteborn, H. P. J. M.; Pedersen, J.; Smith, M. Unintended effects and their detection in genetically modified crops. *Food Chem. Toxicol.* **2004**, *42*, 1089–1125.

(22) Ku, K. M.; Choi, J. N.; Kim, J.; Kim, J. K.; Yoo, L. G.; Lee, S. J.; Hong, Y.-S.; Lee, C. H. Metabolomics analysis reveals the compositional differences of shade grown tea (*Camellia sinensis* L.). *J. Agric. Food Chem.* **2010**, *58*, 418–426.

(23) Louw, L.; Roux, K.; Tredoux, A.; Tomic, O.; Naes, T.; Nieuwoudt, H. H.; van Rensburg, P. Characterization of selected South African young cultivar wines using FTMIR spectroscopy, gas chromatography, and multivariate data analysis. *J. Agric. Food Chem.* **2009**, *57*, 2623–2632.

(24) Stewart, D.; Shepherd, L. V. T.; Hall, R. D.; Fraser, P. D. Crops and tasty, nutritious food – how can metabolomics help? In *Annual Plant Reviews Vol. 43: Biology of Plant Metabolomics*; Hall, R. D., Ed.; Wiley-Blackwell: Oxford, U.K., 2011; pp 181–217.

(25) Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E. Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* **2002**, *1*, 153–161.

(26) Brown, P. N.; Finley, J. P.; Le, C. H.; Murch, S. J. Metabolomic profiling of cranberry (*Vaccinium macrocarpon* Aiton) fruit and cranberry dietary supplements by time-of-flight mass spectrometry with chemometric analysis. *Anal. Bioanal. Chem.* **2011** (in review).

(27) Brown, P. N.; Turi, C. E.; Shipley, P. R.; Murch, S. J. Phytochemical discovery in large cranberry (*Vaccinium macrocarpon* Aiton.), small cranberry (*Vaccinium oxycoccus*) and lingonberry (*Vaccinium vitis-idaea*). *Planta Med.* **2011**, inpress.

(28) Enot, D. P.; Draper, J. Statistical measures for validating plant genotype similarity assessments following multivariate analysis of metabolome fingerprint data. *Metabolomics* **2007**, *3*, 349–355.

(29) Broadhurst, D. I.; Kell, D. B. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2006**, *2*, 171–196.

(30) Tuszynski, J. CRAN – Package caTools, http://cran.r-project.org/web/packages/caTools/index.html (accessed June 2011).

(31) Hand, D. J.; Till, R., J. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. Learn.* **2001**, *45*, 171–186.

(32) Kruskal, W. H.; Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621.

(33) Tusher, V. G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5116–5121.

(34) Schwender, H. Siggenes, http://www.bioconductor.org/packages/2.4/bioc/html/siggenes.html (accessed June 2011).

(35) Eissa, A. B.; Batson, G.; Wise, B. M.; Gallagher, N. B. The process chemometrics approach to process monitoring and fault detection. *J. Process Control* **1996**, *6*, 329–348.

(36) Borowksa, E. J.; Mazur, B.; Kopciuch, R. G.; Buszewki, B. Polyphenol, anthocyanin and resveratrol mass fractions and

antioxidant properties of cranberry cultivars. *Food Technol. Biotechnol.* **2009**, *47*, 56−61.

(37) Frost, C. J.; Mescher, M. C.; Carlson, J. E.; De Moraes, C. M. Plant defense priming against herbivores: getting ready for a different battle. *Plant Physiol.* **2008**, *146*, 818−824.

(38) Vorsa, N.; Polashock, J. J. Alteration of anthocyanin glycosylation in cranberry through interspecific hybridization. *J. Am. Soc. Hortic. Sci.* **2005**, *130*, 711−715.

(39) Varmuza, K.; Filzmoser, P. *Introduction to Multivariate Statistical Analysis in Chemometrics*; CRC Press: Boca Raton, FL, 2009.