JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

J. Agric. Food Chem. 2009, 57, 8705–8710 8705
DOI:10.1021/jf900852y

Variance in the Chemical Composition of Dry Beans Determined from UV Spectral Fingerprints

JAMES M. HARNLY,† MARCIAL A. PASTOR-CORRALES,‡ AND DEVANAND L. LUTHRIA*,†

†Food Composition and Methods Development Laboratory, Beltsville Human Nutrition Research Center and ‡Vegetable Laboratory, Plant Sciences Institute, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland 20705-3000

Nine varieties of dry beans representing five market classes were grown in three locations (Maryland, Michigan, and Nebraska), and subsamples were collected for each variety (row composites from each plot). Aqueous methanol extracts of ground beans were analyzed in triplicate by UV spectrophotometry. Analysis of variance−principal component analysis was used to quantify the relative variance arising from location, variety, between rows of plants, and analytical uncertainty and to test the significance of differences in the chemical composition. Statistically significant differences were observed between all three locations, between all nine varieties, and between rows for each variety. PCA score plots placed the nine varieties in four categories that corresponded with known taxonomic groupings: (1) black beans (cv. Jaguar and cv. T-39), (2) pinto beans (cv. Buster and cv. Othello), (3) small red beans (cv. Merlot), and (4) great northern (cv. Matterhorn and cv. Weihing) and navy (cv. Seahawk and cv. Vista) beans. The relative plant-to-plant variance, estimated from the between row variance, was 71−79% for 25−40 plants per row.

KEYWORDS: Nine common beans; *Phaseolus vulgaris* L.; spectral fingerprinting; multiple locations; analysis of variance; principal component analysis; ANOVA- PCA; UV spectrometry

INTRODUCTION

The chemical composition of dry beans is determined by genetic, environmental, and processing factors. Some genetic factors are obvious to the consumer; pinto beans are readily distinguished from navy beans and black beans. However, the influence of the growing location, seasonal variation (e.g., rainfall, temperature, and total sun exposure), cultivation practices (organic vs conventional farming), and variation between plants can only be determined through a statistical analysis of their chemical compositions.

For nutritional purposes, the nutrient levels and variation are of primary importance. Regretfully, analysis of all of the specific vitamins and minerals can be prohibitively expensive and time-consuming, especially when large degrees of variation are experienced between plants, growing locations, and environmental conditions. Analysis of all of the specific vitamins and trace metals, however, would be time-consuming and costly. Analysis of non-nutrient but bioactive chemical components would further contribute to the cost of characterizing plant materials. A simpler approach is to compare the overall chemical composition of plant materials using spectral fingerprinting or chromatographic profiling. Thus, a single well-characterized plant material can be rapidly compared to new plants from the latest genetic cross, cultivation practice, or processing method.

Spectral fingerprinting is based on direct analysis (no separation) of a sample extract using ultraviolet (UV) and visible (vis) absorption, mass (MS), or nuclear magnetic resonance (NMR) spectrometry or analysis of the solid material using infrared (IR) or near-infrared (NIR) spectrometry (*1*−*7*). Chromatographic profiling employs a separation of the plant extract (or volatile components) by gas (GC) or liquid chromatography (LC) or gel or capillary electrophoresis (CE) with, most commonly, UV, fluorescence (F), or MS detection (*1*−*7*). In both cases, the comprehensiveness of the comparison is dependent on the extraction solvent and procedure that is used. In both cases, an integrated analysis of the chemical composition of the samples requires the use of pattern recognition programs.

Spectral fingerprints, regardless of the means of acquisition, are highly complex, representing the sum of the spectra of each compound present in a sample. In general, it is very difficult to identify, let alone quantify, individual compounds. While identification is sometimes attempted with MS fingerprints, the results are unreliable and chromatographic separation is required to obtain accurate results.

Principal components analysis (PCA) is the most commonly used pattern recognition program for unstructured analysis (*8*). Recently, Harrington et al. (*9*, *10*) reported on the use of analysis of variance (ANOVA)-PCA as a means of isolating experimental factors prior to PCA. This method constructs submatrices of the data for each factor that can be more easily interpreted, visually and statistically, by PCA. Harnly's group (*11*, *12*) reported a variation of ANOVA-PCA that uses the submatrices to compute the relative variance contributed by each factor.

Recent studies have compared the use of UV, vis, and MS spectra of aqueous methanol (60% MeOH and 40% H_2O)

*To whom correspondence should be addressed. Tel: 301-504-7247. Fax: 301-504-8314. E-mail: d.luthria@ars.usda.gov.

Published on Web 09/04/2009

extracts and IR and NIR spectra of solid powders for identifying chemical patterns in broccoli using ANOVA-PCA (*11*−*13*). Composite samples of cultivar (cv.) Legacy grown with four levels of selenium (Se) fertilization and of cv. Majestic grown organically and conventionally, the latter with full irrigation and 80% irrigation (based on transpiration rates), were analyzed. Results for IR, positive ionization (PI), and negative ionization (NI) MS, NIR, and UV showed that between 20 and 40% of the sample variance was due to variety, 50−70% was due to sample treatment (fertilization/irrigation), and 2−15% was due to analytical uncertainty. The only major difference in this pattern was observed using vis spectrometry, which showed 70% of the variance coming from the varieties, 28% from the sample treatment, and 2.2% from analytical uncertainty.

Considering the different physical properties and, hence, the different compounds measured by the different methods, the extent of the agreement of results for IR, NI-MS, PI-MS, NIR, and UV is remarkable. These data suggest that the treatments affected the entire semipolar metabolome (obtained with an aqueous methanol extraction) of the plant. Spectrophotometry in the visible region is less informative since far fewer compounds of the semipolar metabolome have chromophores in this region. Far fewer compounds absorb in the visible portion of the molecular spectrum than in the ultraviolet region. The MS data revealed that differentiation between varieties and treatments was primarily due to differences in the concentration of amino acids, organic acids, and saccharides. The results of these studies point out the need for more research regarding sources of variance in plants.

A recent analytical study characterized the phenolic content of 17 varieties of dry beans (*14*) using a standardized method based on liquid chromatography coupled to a diode array detector and a mass spectrometer (LC-DAD-MS) with positive and negative ionization and low and high (for in-source collision induced dissociation) fragmentation voltages (*15*). Although the beans contained the same hydroxycinnamic acids, the flavonoid components showed distinctive differences. The data suggested that the dry beans could be divided into six categories: (1) black beans, (2) pinto beans, (3) light red kidney beans, (4) pink and dark red kidney beans, (5) small red beans, and (6) alubia, cranberry, great northern, and navy beans.

Thompson et al. (*15*) reported that dry beans in the diet reduced the incidence of breast cancer. The risk factor was dependent on market class (red and white kidney, great northern, small red, navy, and black beans) and growing site (South America and Middle America). Total phenolics and flavonoid contents were strongly associated with the seed coat color. The authors reported the need for profiling beans using metabolomics to identify secondary metabolites responsible for the reduced cancer risk.

In the current study, the UV spectra of nine varieties of beans, comprising four of the categories listed above, were collected and analyzed by ANOVA-PCA. The beans represented a unique set of samples: The nine varieties of beans were each grown in three locations (Maryland, Michigan, and Nebraska), and composite samples were collected for each of the eight rows of each variety at each location. Each sample was prepared and analyzed in triplicate. This paper reports the variance arising from location, variety, plant-to-plant variation, and analytical uncertainty for this group of beans.

MATERIALS AND METHODS

Plant Materials. Nine varieties of beans (Buster, Jaguar, Matterhorn, Othello, Merlot, Seahawk, T-39, Vista, and Weihing) were grown in three locations (Maryland, Michigan, and Nebraska) (*14*). Eight rows of each bean variety were grown in each state with 25−40 plants per row. Composite samples from each row were collected and sent to the lab. The three locations, nine varieties, and eight rows produced a total of 216 samples. Each sample was extracted and analyzed in triplicate.

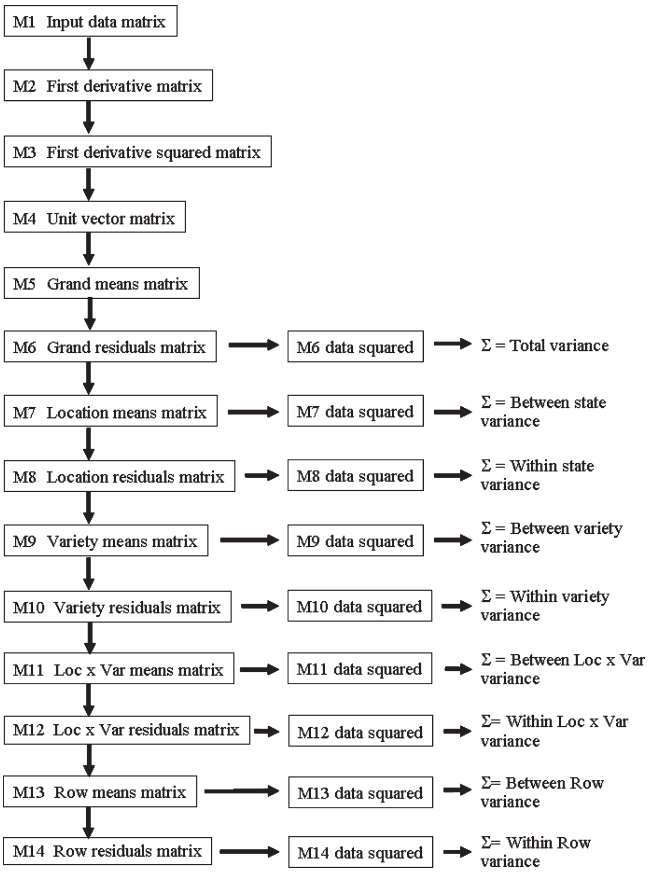

Figure 1. Overview of ANOVA preprocessing.

Chemicals. High-performance liquid chromatography (HPLC)-grade MeOH was purchased from Fisher Chemicals (Fair Lawn, NJ). HPLC-grade acetone was purchased from Burdick & Jackson (Muskegon, MI). Deionized water (18.2 MΩ cm) was obtained in-house using a Nanopure diamond analytical ultrapure water purification system (model D11901, Branstead Internationals, Dubuque, IA). Polyvinylidene difluoride (PVDF) syringe filters with a pore size of 0.45 μm were procured from National Scientific Co. (Duluth, GA).

Extractions. The extraction process was described in detail previously (*11*, *12*, *16*). Briefly, 1 g of weighed freeze-dried and powdered bean samples was extracted three times with 5, 2.5, and 2.5 mL of MeOH:H_2O (60:40, % v/v). The supernates were combined and brought to a final volume of 10 mL with MeOH:H_2O (60:40, % v/v). All extracts were stored in 2 mL HPLC vials under nitrogen at −70 °C until analyzed. Prior to analysis, extracts were filtered and diluted by a factor of 50 into MeOH:H_2O (60:40, % v/v).

UV Instrumentation. UV spectral fingerprints of the bean extracts were acquired as previously described (*11*). All data were recorded on a Lambda 25 spectrophotometer (Perkin-Elmer, Boston, MA). UV fingerprints were acquired at 1 nm intervals for the region 200−400 nm. The extracts were diluted by a factor of 50 due to the strong absorbance in this region.

Data Analysis. All spectral data were exported from the UV spectrophotometer to Excel (Microsoft, Inc., Belleview, WA) for preprocessing. PCA was performed using Pirouette 3.1 (Infometrix, Inc., Bothell, WA) and Solo (eigenvector, Wenatchee, WA).

Individual profiles were combined to produce a two-dimensional (sample × wavelength) input data matrix (M1 in **Figure 1**) for ANOVA preprocessing. The data were first transformed to the first derivative using Savitsky−Golay functions (M2). The ANOVA preprocessing, which is

Table 1. Flavonoids and Phenolic Acids Found in Beans[a]

category[a]	market class	variety	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
			peak nos.[a]																																								
			flavonols												anthocyanins								anthcyndns					hydroxycinnamic acids/acid derivatives															
1	black	Jaguar	E				E	E	E	E	E				E	E	E											E	E	E		E	E	E	E		E				E	E	E
		T-39	E				E	E	E	E	E				E	E	E											E	E	E		E	E	E	E		E				E	E	E
2	pinto	Buster							E		E	E		E														E	E	E		E	E	E	E		E				E	E	E
		Othello							E		E	E																E	E	E		E	E	E	E		E				E	E	E
3	small red	Merlot							E		E	E	E/H					E	E	E	E	E				H	H	E	E	E		E	E	E	E		E	H	H	H	E	E	E
4	great northern	Matterhorn																										E	E	E		E	E	E	E		E				E	E	E
		Weihing																										E	E	E		E	E	E	E		E				E	E	E
	navy	Seahawk																										E	E	E		E	E	E	E		E				E	E	E
		Vista																										E	E	E		E	E	E	E		E				E	E	E

[a] Category and peak numbers refer to peaks in chromatograms presented in ref XX. E, found in estract; H, found in hydrolized estract; and NA, not analyzed in this study.

based on the method of Harrington et al. (*9*), has been described in detail previously (*11*, *12*) and is shown in **Figure 1**. All data in the first derivative matrix (M2) were squared (M3), and the sums for individual spectra were used to create a unit vector matrix (M4). All data were then mean centered for each measured variable to give the grand means matrix (M5) and subtracted from the unit vector matrix to provide the grand residuals matrix (M6 = M4 − M5). The squares of the grand means residuals were used to compute the total variance. The grand means residuals (M6) were used to compute the location means matrix (M7) and the location residuals matrix (M8 = M6 − M7). The squares of the data in M7 and M8 were used to compute the variance between and within the location, respectively. In a similar manner, the location residuals (M8) were used to compute the variety means (M9) and the variety residuals (M10 = M8 − M9). The squares of M9 and M10 were used to compute the variance between and within the variety, respectively. The variety residuals (M10) were used to compute the means (M11) and residuals (M12) matrices arising from location−variety interaction (loc × var). Whereas the location and variety means matrices each consisted of two sets of averages, the loc × var means matrices consisted of four sets of averages. Squares of M11 and M12 were used to compute the variance between and within loc × var, respectively. The loc × var residuals were used to compute the individual row means (M13) and the row residuals (M14 = M12 − M13). The squares of M11 and M12 were used to compute the variance between and within the rows, respectively.

The loc × var residuals matrix (M12) represents the biological (genetic and environmental) and analytical uncertainty. The row residual matrix (M14) represents the analytical uncertainty arising from the three repeat analyses of each sample. The biological and analytical uncertainty matrix (M12) was added to the location means matrix (M7) and the variety means matrix (M9) to generate the location and variety test matrices that were submitted to PCA.

Statistical Calculations. The significance of the separation of two populations was computed using Student's *t* test. Student's *t* value and the combined standard deviation of the means were computed in the standard manner (*17*).

RESULTS AND DISCUSSION

ANOVA-PCA. Samples of 17 varieties of beans, representing 10 market classes, were previously analyzed using our standardized profiling method for phenolic compounds (*14*). The 17 varieties were divided into six categories based on visual inspection of the data. As shown in **Table 1**, the nine varieties grown in the present study come from five market classes and fall into four of the previously determined categories. In general, **Table 1** shows that the presence or absence of specific phenolic compounds is correlated with the categories. Two of the market classes (great northern and navy) form one category because of the similarity of their phenolic content. The earlier data were not subjected to pattern recognition analysis.

Using the ANOVA-PCA method described by Harrington et al. (*9*), the data matrix consisting of the UV spectra for all of the bean samples was used to establish a series of submatrices (M6−M12 in **Figure 1**) that were used for PCA and computing variance. Matrices subjected to PCA were obtained by adding the "biological uncertainty" matrix (M12) to the location and variety means matrices (M7 and M9, respectively).

Harnly's group (*11*, *12*) showed that variance for each of the experimental factors (**Table 2**) could be computed by summing the squares of the individual data in each matrix (M6−M14 in **Figure 1**). Thus, the total variance (M6) can be separated into the variance between (M7) and within locations (M8). Similarly, the within location variance (M8) can be separated into the variance between (M9) and within (M10) variety. The within variety variance (M10) can be broken down into between (M11) and within (M12) location−variety interaction (loc × var) variance. The within loc × var variance can be separated into the variance between (M13) and within (M14) rows. The within loc × var variance (M12) and the within row variance (M14) are also known as the biological and analytical uncertainty variance, respectively.

The variance due to location−variety interaction is computed prior to the variance between and within rows because location and variety are regarded as fixed factors; that is, over time, their effects would be expected to be nonrandom or predictable. Rows, on the contrary, are random factors. There is no reason that, for a specific cultivar, row 1 in Nebraska would correlate with row 1 in Maryland.

Table 2 shows that 2.4% of the total variance arose from location (M7), 71.2% from variety (M9), 11.0% from loc × var (M11), and 15.4% from biological uncertainty (M12). The biological uncertainty was attributed to 8.8% from row-to-row variance and 6.6% from analytical uncertainty. Thus, the variance due to variety exceeds the other sources of variance. **Figure 2** shows a plot of the variance between locations, between variety, and from biological uncertainty. Because of the greater variance due to variety, variance from location and biological uncertainty has been multiplied by a factor of 10 to make all of the plots equally visible.

The data in **Table 2** can be expressed as a classic two-way crossed ANOVA with eight samples in each category (**Table 3**) (*17*−*19*) since all nine varieties were grown at all three locations and a composite sample was taken from each of the eight rows of plants. Location and variety are fixed factors, and row is a random factor (*18*, *19*). The high significance of the loc × var *F* value shows that there is a low probability that the variety

Table 2. Variance from ANOVA Preprocessing (**Figure** 1)

source of variance	location	variety	loc × var	row	% variance
total (M6)	11.935				100.0
between location (M7)	0.292				2.4
within location (M8)	11.643				
between variety (M9)		8.495			71.2
within variety (M10)		3.148			
between (location × variety)(M11)			1.314		11.0
within (location × variety) (M12) (biological uncertainty)			1.834		15.4
between row (M13)				1.049	(8.8)
within row (M14) (analytical uncertainty)				0.784	(6.6)

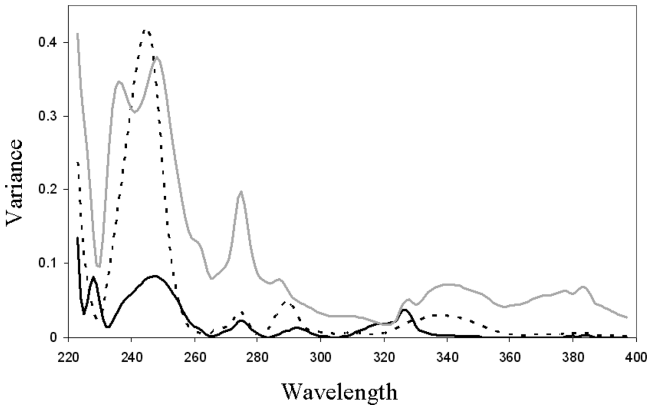

Figure 2. Variance as a function of wavelength for location × 10 (solid black), variety × 1 (dashed black), and biological uncertainty × 10 (solid gray).

means are consistent relative to one another at all three locations. Thus, some of the variety means differ statistically among the three locations, that is, the relative means of the nine varieties are location-dependent.

Location Variance. The previous section showed that the variance due to location was 2.4% of the total variance (**Table 2**) and that there was < 0.0001 probability that the means were the same (**Table 3**). **Figure 3** shows plots of the PCA scores obtained using the location test matrix (M7 + M12) for Michigan vs Nebraska (MI vs NE) and for Maryland vs Nebraska (MD vs NE). The plot of Maryland and Michigan (MD vs MI) (not shown) was similar to MD vs NE (**Figure 3B**). Each score plot is composed of 216 samples from each state representing all nine varieties, eight subsamples, and triplicate analyses.

Table 4 presents the means and standard deviations for the horizontal distribution of the data for each of the PCA plots in **Figure 3** (MI vs NE and MD vs NE) and for the plot not shown (MD vs MI). Visual inspection of the plots in **Figure 3** suggests that location had little effect on the chemical composition of the beans. However, as shown in **Table 4**, the combined standard deviations of the means are dramatically reduced as a result of the large number of sample and the t values are statistically significant. The probabilities of the means being were < 0.0001. These probabilities are consistent with those from the two-way ANOVA in **Table 3**. Both the two-way ANOVA (**Table 3**) and the ANOVA-PCA (**Table 4**) display the same relationship, although, in the latter case, the data have been transformed by PCA. In both cases, the biological uncertainty is the limiting variance for the F values (**Table 3**) and t values (**Table 4**).

Table 3. Crossed ANOVA

source	DF	sum of squares	mean square	F value	P
between location	2	0.292	0.146	15.0	<0.0001
between varietiy	8	8.49	1.06	109.0	<0.0001
between (location × variety)	16	1.31	0.082	8.45	<0.0001
within (location × variety) (biological uncertainty	189	1.834	0.0097		

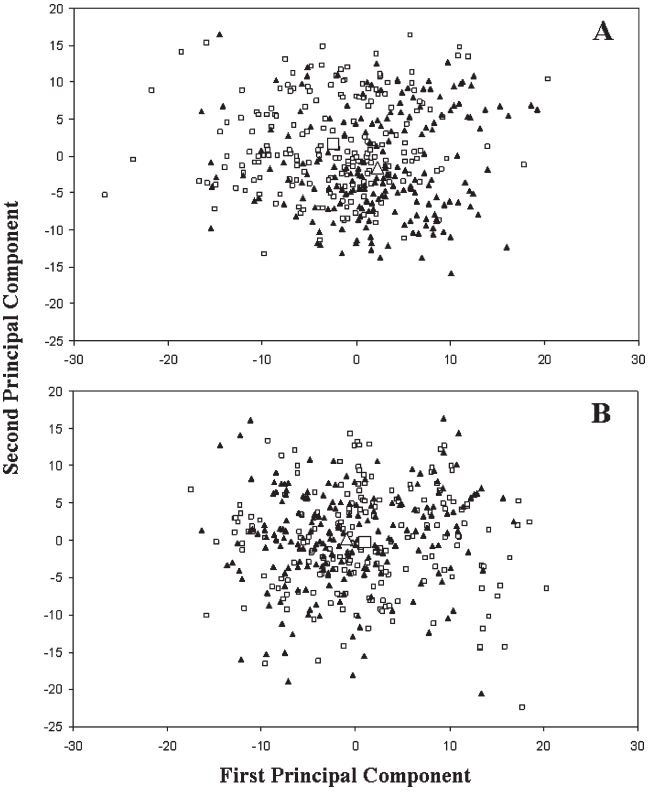

Figure 3. PCA score plots from location test matrix (M7 + M12) comparing beans from three locations: (**A**) Michigan (□) and Nebraska (■) and (**B**) Maryland (□) and Nebraska (■). Large crossed symbols show the center of mass.

Table 4. Statistical Significance of Distance between Location Means (**Figure** 3)[a]

	MD vs MI	MD vs NE	MI vs NE
no. of data points	432	432	432
distance between means	2.56	1.92	4.70
standard deviation of the mean	0.72	0.72	0.69
Student's t value	3.55	2.65	6.84
probability	<0.0001	<0.0001	<0.0001
$(X_{ave,1} - X_{ave,2})/\sigma_{ind}$ (223−397 nm)	0.24	0.18	0.46
$(X_{ave,1} - X_{ave,2})/\sigma_{ind}$ (315−328 nm)	0.37	0.39	0.80

[a] MD, Maryland; MI, Michigan; and NE, Nebraska.

The large overlap of the data clusters in **Figure 3** indicates that it would be difficult to develop a statistical model to predict the origin of individual samples. Complete separation of the clusters would be necessary to ensure accurate prediction. Consequently, computation of a t value based on the individual standard

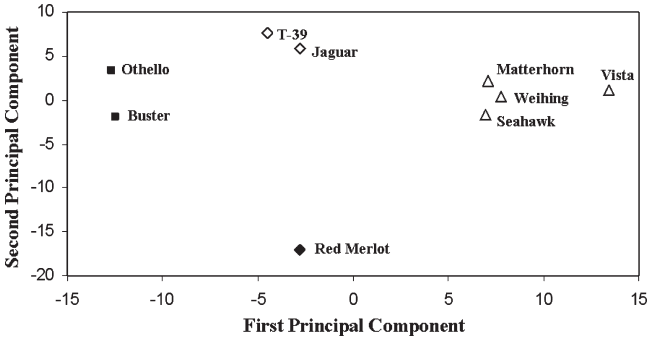

Figure 4. PCA scores of the variety means matrix (M9) comparing the nine bean varieties in four groupings: black beans (Jaguar and T-39) (◇), pinto beans (Buster and Othello) (■), small red beans (Merlot) (♦), and navy and great northern beans (Matterhorn, Seahawk, Vista, and Weihing) (△).

deviation, $X_{ave,1}-X_{ave,2})/\sigma_{ind}$, would be more predictive (**Table 4**). An alternate approach is to select a smaller wavelength range where the variance from location is greater as compared to the biological uncertainty. In **Figure 2**, the two variances are approximately the same between 315 and 328 nm. Use of this region for ANOVA-PCA provided improved (by a factor of 2), but not complete, separation as shown in **Table 4**.

From a consumer's point of view, a very large number of servings of beans from a specific location would have to be eaten before an individual would be exposed to a significantly different chemical intake. The possibility of this occurring would be low since the beans available in grocery stores come from a wide variety of locations that are not identified to the consumer. In addition, it is not known if the chemical difference of the plants, arising from different weather conditions and soil chemistry of the different location, between locations would have biological significance.

Variety Variance. Table 2 shows that 71% of the total variance was due to variety, and **Table 3** shows that the probability was <0.0001 that the means were the same. The data in **Table 1** suggest that the nine varieties should fall into four categories. **Figure 4** presents the score plot obtained from PCA of the varieties means matrix (M9). Because the analytical uncertainty is omitted, each variety shows up as a single point. **Figure 4** shows that the nine bean varieties fell into four clusters that corresponded to the four categories that were previously identified (*14*). These data confirm that the color differences of the beans correspond to chemical differences besides the chromaphores. The four groups are composed of the black beans (cv. Jaguar and cv. T-39), pinto beans (cv. Buster and cv. Othello), small red beans (cv. Merlot), and white beans. The latter composed of the larger great northern beans (cv. Matterhorn and cv. Weihing) and the smaller navy beans (cv. Seahawk and cv. Vista).

A comparison of three of the bean varieties is shown in the PCA score plots in **Figure 5**. In **Figure 5A**, cv. Jaguar is compared to cv. T-39, a comparison within the same category, and in **Figure 5B**, cv. Jaguar is compared to cv. Merlot, a comparison between categories. Each comparison is based on 72 samples (eight rows of each variety from three locations analyzed in triplicate). The t values for the two comparisons are 3.7 and 41.7, respectively, and the probability that the means are equal is <0.0001 in both cases. This is visually obvious in **Figure 5B** and less intuitive in **Figure 5A**. These probabilities are consistent with those from the nested two-way ANOVA in **Table 3**.

A comparison of all nine varieties at each location (data not shown) yielded t values that showed a probability of <0.0001

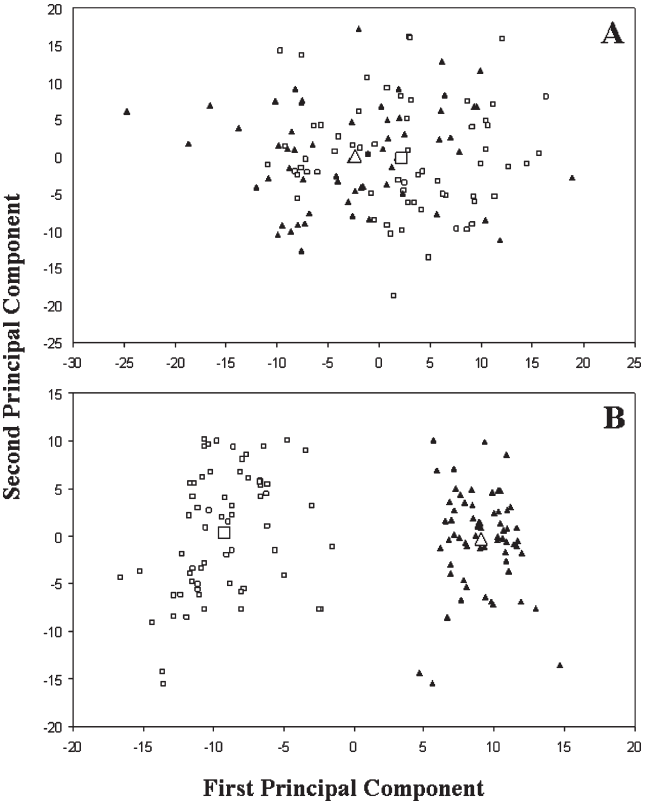

Figure 5. PCA scores of the variety means test matrix (M9 + M12) comparing pairs of bean varieties within groupings (**A**) black beans cv. Jaguar (■) and cv. T-39 (□) and between groupings (**B**) black beans cv. Jaguar (■) and small red beans cv. Merlot (□). Large crossed symbols show the center of mass.

that the means are the same. In general, the t values for comparisons within a category were less than those for comparisons between categories. The score plots in **Figure 5** represent the two extremes; the t value for cv. Jaguar and cv. T-39 were the smallest and the t value for cv. Jaguar and cv. Merlot were the largest for all variety comparisons in all three locations. All other comparisons had t values falling somewhere between these two.

The comparisons of the mean values in the preceding paragraphs were based on a t test that employed the standard deviation of the mean, that is, the combined standard deviation of the individual clusters divided by the square root of the number of samples. Consequently, the means of highly overlapping clusters (**Figure 5A**) can be statistically different. However, it is intuitive that the possibility of establishing a model to predict the variety of individual samples would be far more likely for Jaguar and T-39 than for Jaguar and Merlot (**Figure 5A**). A more accurate predictor for successful modeling is visual confirmation that the clusters are separated or a t value based on the individual standard deviation. In the latter case, the revised t values ranged from 0.5 to 7.7. In general, it would be possible to distinguish between individual samples from different groupings but not within groupings (**Figure 4**).

Plant-to-Plant Variance. Table 2 shows that approximately 15% of the total variance came from the variation between rows, while **Table 3** shows that the probability is <0.0001 that the row means are the same. The score plots for the comparisons of individual rows showed six points (each row sample extracted and analyzed in triplicate) with varying degrees of overlap (data

Table 5. Plant-to-Plant Variance

	original data		25 plants/row		40 plants/row	
source of variance	variance	% total	variance	% total	variance	% total
between location	0.29	2.4	0.29	0.8	0.29	0.6
between variety	8.50	71.2	8.50	22.9	8.50	16.1
between location × variety	1.31	11.0	1.31	3.5	1.31	2.5
between row	1.05	8.8				
between plant			26.2[b]	70.7	42.0[c]	79.4
analytical uncertainty	0.78	6.6	0.78	2.1	0.78	1.5
total	11.9[a]	100.0	37.1[a]	100.0	52.9[a]	100.0

[a] Sum of variances. [b] 1.05 × 25. [c] 1.05 × 40.

not shown). An example is the comparison of all 28 possible combinations of the rows of Merlot in Nebraska. The t values ranged from 0.25 to 10.02, a typical range for all the samples. Of the 28 t values, 13 were statistically significant at the 95% confidence level, 12 were significant at the 80% level, and three were significant at the 50% level.

Plant-to-plant variance is much more interesting than row-to-row variance. In this study, the beans from each row are a composite of 25−40 plants. The plant-to-plant variance should therefore be 25−40 times greater than the row-to-row variance, since the variance of the mean is equal to the individual variance divided by n, the number of samples. Calculation of the plant-to-plant variance is presented in **Table 5** based on the original experimental data in **Table 3**. For 25 plants per row, the plant-to-plant variance is 26.2 (= 25 × 1.05). The total variance was then determined as the sum of all of the variances. Thus, for 25 and 40 plants per row, the relative variance between plants was 71 and 79%, respectively.

This rough approximation suggests that for an individual plant, the plant-to-plant variance is at least seven times greater than that due to variety. Eight, or more, plants must be composited before the variance due to variety becomes equal or dominant. It should be remembered that the data in **Table 5** are expressed as the percentage of total variance. These values for plant-to-plant variance cannot be equated to the relative variance for individual components (i.e., $\sigma^2/X_{\text{average}}$). As the old saying goes, this would be comparing apples and oranges.

CONCLUSION

UV spectrophotometry, as evaluated here, is surprisingly sensitive, allowing discrimination between beans with respect to location, variety, and between plants based on their chemical composition. These results are consistent with those previously reported for broccoli. The detectable chemical differences suggest that there may also be a difference in the nutritional quality. Although bean varieties are readily discerned by visual inspection, the variance due to location and between plants can only be determined experimentally. The variance contributed by each of these experimental factors in, descending order, was between plants > variety > location. The rapid method of comparison reported here is considerably simpler and cheaper than full plant nutrient profiles.

ACKNOWLEDGMENT

We appreciate helpful discussions with Dr. Bryan Vinyard, Agriculture Research Service, U.S. Department of Agriculture, and Prof. Peter Harrington, Ohio University.

LITERATURE CITED

(1) Sumner, L. W.; Mendes, P.; Dixon, A. Plant metabolomics: Large-scale phytochemistry in the functional genomics era. *Phytochemistry* **2003**, *62*, 817–836.
(2) Goodacre, R.; York, E. V.; Heald, J. K.; Scott, I. M. Chemometric discrimination of unfractionated plant extracts analyzed by electrospray mass spectrometry. *Phytochemistry* **2003**, *62*, 859–863.
(3) Dunn, W. B.; Overy, S.; Quick, W. P. Evaluation of automated electrospray-TOF mass spectrometry for metabolic fingerprinting of the plant metabolome. *Metabolomics* **2005**, *1*, 137–148. Goodacre, R.; Luned, R.; Ellis, D. I.; Thorogood, D.; Reader, S. M.; Ougham, H.; King, I. From phenotype to genotype: Whole tissue profiling for plant breeding. *Metabolomics* **2007**, *3*, 489–501.
(4) Dunn, W. B; Ellis, D. I. Metabolomics: Current analytical platforms and methodologies. *Trends Anal. Chem.* **2005**, *24*, 285–294.
(5) Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. Metabolite fingerprinting and profiling in plants using NMR. *J. Exp. Biol.* **2005**, *56*, 255–265.
(6) Mattoli, L.; Cangi, F.; Maidecchi, A.; Ghiara, C.; Ragazzi, E.; Tubaro, M.; Stella, L.; Tisato, F.; Traldi, P. Metabolomic fingerprinting of plant extracts. *J. Mass Spectrom.* **2006**, *41*, 1534–1545.
(7) Shulaev, V.; Cortes, D.; Miller, G.; Mittler, R. Metabolomics for plant stress response. *Physiol. Plant.* **2008**, *132*, 199–208.
(8) Wold, S. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52.
(9) Harrington, D. B.; Vieira, N. E.; Espinoza, J.; Kien, J. K.; Romero, R.; Yergey, A. L. Analysis of variance-principal component analysis: A soft tool for proteomic discovery. *Anal. Chim. Acta* **2005**, *544*, 118–127.
(10) Harrington, D. B.; Vieira, N. E.; Ping, C.; Espinoza, J.; Kien, J. K.; Romero, R.; Yergey, A. L. Proteomic analysis of amniotic fluids using analysis of variance-principal component analysis and fuzzy rule-building expert systems applied to matrix-assisted laser desorption/ionization mass spectrometry. *Chemom. Intell. Lab. Syst.* **2006**, *82*, 283–293.
(11) Luthria, D. L.; Mukhopadhyay, S.; Robbins, R. J.; Finley, J. W.; Banuelos, G. S.; Harnly, J. M. UV spectral fingerprinting and analysis of variance-principal component analysis: A useful tool for characterizing sources of variance in plant materials. *J. Agric. Food Chem.* **2008**, *56*, 5457–5462.
(12) Luthria, D. L.; Mukhopadhyay, S.; Robbins, R. J.; Finley, J. W.; Banuelos, G. S.; Harnly, J. M. Discriminating between cultivars and treatments of broccoli using mass spectral fingerprinting and analysis of variance principal component analysis. *J. Agric. Food Chem.* **2008**, *56*, 9819–9827.
(13) Luthria, D. L.; Mukhopadhyay, S.; Harnly, J. M. A comparison of methods for spectral fingerprinting. Manuscript in preparation.
(14) Lin, L.-Z.; Harnly, J. M.; Pastor-Corrales, M. S.; Luthria, D. L. Polyphenolic profiles of common bean (*Phaseolus vulgaris* L.). *Food Chem.* **2008**, *107*, 399–410.
(15) Thompson, M. D.; Brick, M. A.; McGinley, J. N.; Thompson, H. J. Chemical composition and mammary cancer inhibitory activity of dry bean. *Crop Sci.* **2009**, *49*, 179–186.
(16) Lin, L.-Z.; Harnly, J. M. A screening method for the identification of glycosylated flavonoids and other phenolic compounds using a standard analytical approach for all plant materials. *J. Agric. Food Chem.* **2007**, *55*, 1084–1095.
(17) Wine, R. L. *Statistics for Scientists and Engineers*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1964.
(18) McDonald, J. H. *Handbook of Biological Statistics*; http://udel.edu/~mcdonald/statintro.html.
(19) Steel, R. G. D. Torrie, J. H. *Principles and Procedures of Statistics—A Biometrical Approach*, 2nd ed.; McGraw-Hill: New York, NY, 1980; p 360.

Received March 12, 2009. Revised manuscript received May 12, 2009. Accepted August 1, 2009. This research was supported by the Agricultural Research Service of the U.S. Department of Agriculture and the Office of Dietary Supplements at the National Institutes of Health.