Approaches to Food Varietal Differentiation Using Accurate-Mass Q-TOF LC/MS with Molecular Feature Extraction, Database Mining, and Principal Component Analysis

Application Note

Food

Authors

Alyson E. Mitchell, Jihyun Lee,
Department of Food Science and Technology, University of California, Davis
One Shields Avenue
Davis, CA 95616

Jerry Zweigenbaum
Agilent Technologies, Inc.
2850 Centerville Road
Wilmington, DE 19808

Abstract

Onions are a major source of quercetin, an important dietary flavonoid. Using Agilent MassHunter PCDL Manager Software, publicly available literature, and online databases such as Phenol-Explorer and ChemSpider, an exact-mass onion-specific flavonoid database was created. The Agilent MassHunter Qualitative Analysis Software was used to search Q-TOF LC/MS data of extracts of seven varieties of onions for matches with the flavonoid database. Nineteen flavonoids were tentatively identified. Based on the flavonoids tentatively identified, a principal component analysis (PCA) of the onion varieties for varietal and color difference was performed using Agilent Mass Profiler Professional Software. The PCA showed separation in varietal and color. For comparison, a PCA of unknown compounds determined by molecular-feature extraction of possible compounds, filtered on uniqueness in the onion varieties, was performed. Though this PCA produced similar results, the top 10 tentatively identified nonanthocyanidin flavonoids were more strongly associated with color and variety.

Agilent Technologies

Introduction

A complementary report of this study of Accurate Mass Q-TOF LC/MS approaches to onion varietal differentiation is published in Physical Methods in Food Analysis, ACS Symposium Series [1].

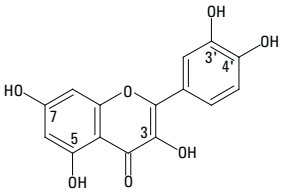
Fruits and vegetables are the primary dietary source of vitamins, minerals, fiber, and a wide array of nonessential nutrient phytochemicals such as polyphenolic antioxidants, flavonoids, carotenoids, alkaloids, and glucosinolates. Many of these compounds exhibit potent biological activity. The range of bioactive phytochemicals expressed are species and cultivar dependent. Bioactive phytochemicals have a wide range of chemical structures. Flavonoids, the most abundant subclass of plant-derived polyphenolic bioactive compounds, can undergo many chemical modifications including acylation, malonylation, sulfation, methylation, and glycosylation. Their glycoside forms can be mono-, di-, and tri-saccharide substituted. The health benefits associated with the consumption of fruits and vegetables are largely thought to be due to the synergistic activities of these bioactive phytochemicals.

Major US dietary sources of quercetin, an important dietary flavonoid, are onions with 210 mg/kg fresh weight, and apples with 30 mg/kg fresh weight. Interestingly, most human and cell culture data is derived from quercetin forms not found in foods (Figure 1).

Genetics are the predominant factor influencing the complement of flavonoids in foods. However, levels can also be influenced by agronomic and environmental factors (for example, soil amendments, growing region, and UV light exposure.)

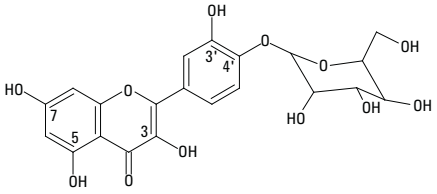
Currently, there is little understanding of how variety, growing season and region, processing, storage, and packaging influence bioactive chemical composition. Likewise, the effects of these factors on the nutritive value of foods are not well understood. This lack of knowledge makes the medicinal or personalized nutritional use of these foods, or manufacturing food-based products with consistent amounts of bioactives, difficult.

This application note used onions as a model to explore how Accurate-Mass Q-TOF LC/MS with Agilent MassHunter Personal Compound Database Library (PCDL) manager, MassHunter Qualitative Analysis, and MassHunter Mass Profiler Professional Software, could be used to identify varietal differences, without standards, using nontargeted and unknown approaches.



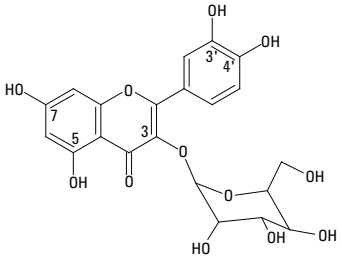
Quercetin aglycone (I)

Used in cell culture studies, clinical trials, and sold as an ingredient.



Quercetin-4'-*O*-β-D-glucoside (II)

Primary form found in onions.



Quercetin-3-*O*-β-D-galactoside (III)

Primary form found in apples.

Figure 1. Examples of quercetin forms.

Experimental

A detailed description of the experimental procedures can be found in the complementary report published Physical Methods in Food Analysis, ACS Symposium Series [1].

Sample preparation

Four varieties of yellow onions (Cowboy, Chief, Vaquero, and Sommerset) and three varieties of red onions (Red Rock, Salsa, and Merenge), obtained from Gills Onion of Oxnard, California, USA, were evaluated.

A general extraction procedure was used to perform an analysis without targeting specific compounds. Because the inner onion layers have limited anthocyanidins (red color), they were separated from the outer layers and lyophilized and extracted with 80% MEOH for 20 minutes. Extraction conditions excluded anthocyanidins from differentiating the red versus the yellow varieties. The extraction was performed in triplicate.

Instrumentation

The extracted samples were analyzed using an Agilent 1290 Infinity LC System coupled to an Agilent 6530 Accurate-Mass Quadrupole Time-of-Flight (Q-TOF) LC/MS system with Agilent Jet Stream technology for electrospray ionization. The 1290 Infinity LC System was equipped with a binary pump with an integrated vacuum degasser (G4220A), an autosampler (G4226A) with a thermostat (G1330B), and a thermostatted column compartment (G1316C). The 1290 Infinity LC System parameters are shown in Table 1.

Table 1. Agilent 1290 Infinity LC System Parameters

Instrument	Agilent 1290 Infinity LC System
Mobile phases	(A) 0.1% formic acid in water, initial 5% B (B) 0.1% formic acid in acetonitrile
Gradient	Linear Time (min) %B 0–5 5–10 5–8 10–12 8–10 12–15 10–15 15 15–18 15–55 18–20 55–90
Flow rate	0.4 mL/min
Column	Agilent Poroshell 120 EC-C18, 2.1 × 100 mm, 2.7 µm (p/n 695775-902)
Post run time	4 minutes at initial mobile phase
Temperature	30 °C
Injection volume	5 µL

A Q-TOF LC/MS analysis of the samples was performed. To identify all possible flavonoids, the instrument methodology was defined as broadly as possible. Total ion spectra were collected over a mass range of *m/z* 100–1,000, in both negative and positive modes. Electrospray ionization generally ionizes a broad range of compounds, with the exception of those that are relatively nonpolar. The flavonoids were expected to respond well, but their relative response was highly impacted by their endogenous chemical structure (for example, conjugation site, type of sugar, acylation, and so forth). Some responded only in positive ion mode and others only in negative mode. Those that responded in both modes may have had much higher ionization efficiency in one mode versus the other. The Q-TOF MS parameters are shown in Table 2.

Table 2. Q-TOF MS Parameters

Instrument	Agilent 6530 Accurate Mass Q-TOF LC/MS
Ionization mode	Positive and negative electrospray with Agilent Jet Stream technology
Acquisition rate	1.0 spectra/s
Mass range	100–1,000 *m/z*
Drying gas temperature	225 °C
Drying gas flow rate	8.0 L/min
Sheath gas temperature	300 °C
Sheath gas flow rate	10.0 L/min
Nebulizer gas	45 psi
Skimmer voltage	65 V
Octopole RF	750 V
Fragmentor	125 V
Capillary	2.5 kV

Continuous infusion of the reference ions *m/z* 119.0363 (proton abstracted purine) and 966.0007 (formate adduct of hexakis (^{1}H, ^{1}H, ^{3}H-tetrafluoropropoxy) phosphazine or HP-921) in negative ion mode was used to correct each spectrum to achieve accurate mass measurement typically better than 2 ppm.

Data analysis

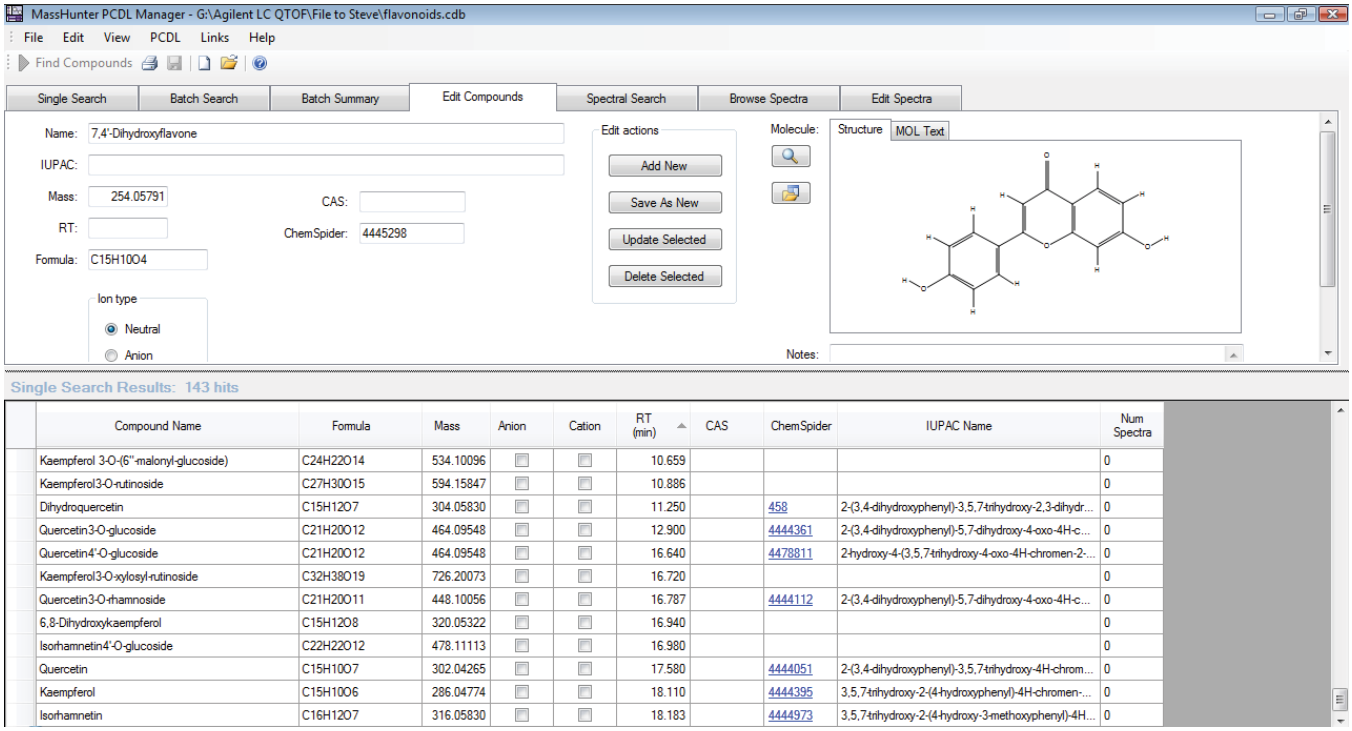
Data analysis was performed in four steps:

1. A flavonoid database was created and customized for the analysis of onions.
2. Data obtained from the Q-TOF LC/MS single MS analyses were searched for matches with the flavonoid database and the matches were scored.
3. Based on the compound matches (nontargeted results, that is without using standards for identification or confirmation), a principal component analysis (PCA) of the onion varieties for varietal and color differences was performed.
4. For comparison, a PCA of all possible compounds found (unknowns, no matching used) in the onion varieties was performed.

Database creation

Using public domain literature, including databases such as Phenol-Explorer (www.phenol–explorer.eu/) and ChemSpider (www.chemspider.com), 150 possible flavonoids, including conjugated forms that could possibly be present in onions, were identified. The molecular formula of each flavonoid was input into the MassHunter PCDL Manager Software (Figure 2) to create a database. Using the molecular formula entered, the MassHunter PCDL Manager Software automatically calculates the exact masses. This database was then used in the MassHunter Qualitative Analysis Software to search the data files obtained from each sample for matches with these compounds.

The MassHunter PCDL Manager Software allows input of compound names, IUPAC names, associated notes, links, structure, and other useful information. If a structure is available in .mol format, the software calculates the molecular formula and exact mass using the imported .mol file. The database component of the MassHunter PCDL Manager Software contains molecules, not ions. The library component provides the ability to add exact mass MS/MS spectra for any of the compounds in the database. Simple searches can be performed using the MassHunter PCDL Manager Software, but searching complete data files requires use of the MassHunter Qualitative Analysis Software.



Compound Name	Formula	Mass	Anion	Cation	RT (min)	CAS	ChemSpider	IUPAC Name	Num Spectra
Kaempferol 3-O-(6''-malonyl-glucoside)	C24H22O14	534.10096			10.659				0
Kaempferol3-O-rutinoside	C27H30O15	594.15847			10.886				0
Dihydroquercetin	C15H12O7	304.05830			11.250		458	2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-2,3-dihydr...	0
Quercetin3-O-glucoside	C21H20O12	464.09548			12.900		4444361	2-(3,4-dihydroxyphenyl)-5,7-dihydroxy-4-oxo-4H-c...	0
Quercetin4'-O-glucoside	C21H20O12	464.09548			16.640		4478811	2-hydroxy-4-(3,5,7-trihydroxy-4-oxo-4H-chromen-2-...	0
Kaempferol3-O-xylosyl-rutinoside	C32H38O19	726.20073			16.720				0
Quercetin3-O-rhamnoside	C21H20O11	448.10056			16.787		4444112	2-(3,4-dihydroxyphenyl)-5,7-dihydroxy-4-oxo-4H-c...	0
6,8-Dihydroxykaempferol	C15H12O8	320.05322			16.940				0
Isorhamnetin4'-O-glucoside	C22H22O12	478.11113			16.980				0
Quercetin	C15H10O7	302.04265			17.580		4444051	2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-4H-chrom...	0
Kaempferol	C15H10O6	286.04774			18.110		4444395	3,5,7-trihydroxy-2-(4-hydroxyphenyl)-4H-chromen-...	0
Isorhamnetin	C16H12O7	316.05830			18.183		4444973	3,5,7-trihydroxy-2-(4-hydroxy-3-methoxyphenyl)-4H...	0

Figure 2. Agilent MassHunter PCDL Manager Software was used to build an exact mass database for the analysis of onions.

Database search

Using MassHunter Qualitative Analysis Software (Figure 3), the data obtained from the Q-TOF LC/MS analysis was searched against the PCDL created in Step 1 of the data analysis. Specifically, using the software's find-by-formula feature, the accurate mass tolerances were set and then applied to a search of the data file for ions of the expected isotopes, adducts (for example, H+, Na+), dimers, trimers, and so on.

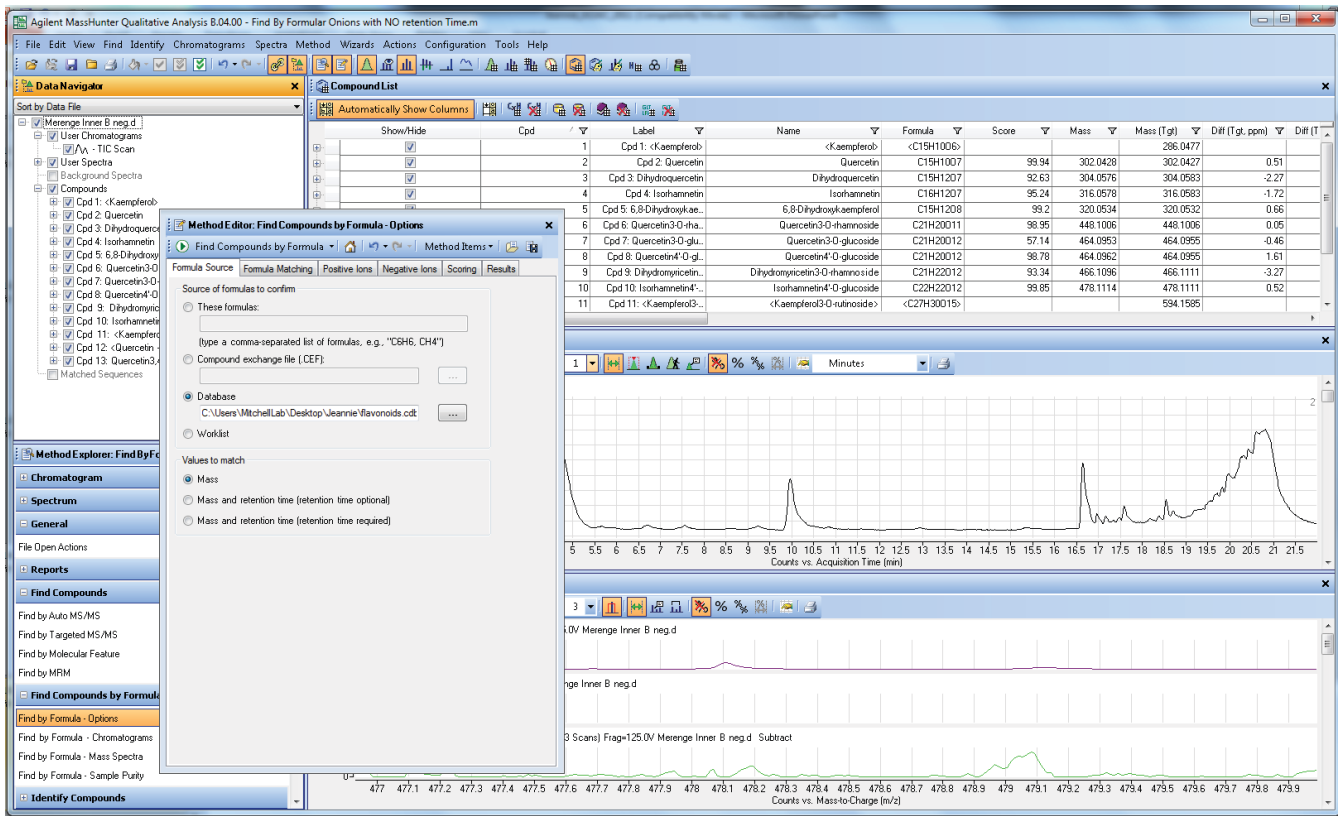


Figure 3. Agilent MassHunter Qualitative Analysis Software was used to search the data for matches with the formulas in the flavonoid database.

The experimentally measured mass of each ion was evaluated against the exact mass of the molecular formula and the expected isotope pattern for that formula. This data was used to score the match of the ions with the database (Figure 4). Using this approach, the flavonoids were tentatively identified by the software without using standards. We refer to this approach as a targeted search for nontarget compounds.

An exact match of mass, isotopic mass, and isotope spacing can only confirm the molecular formula, and not whether it is actually the compound in the database. When standards are not used, MS/MS can provide additional information that can be used to make the identification. If standards are available, retention times and MS/MS can also be added to the database and used to score the match. Confirmation requires comparison to a standard analyzed using the same conditions as the sample.

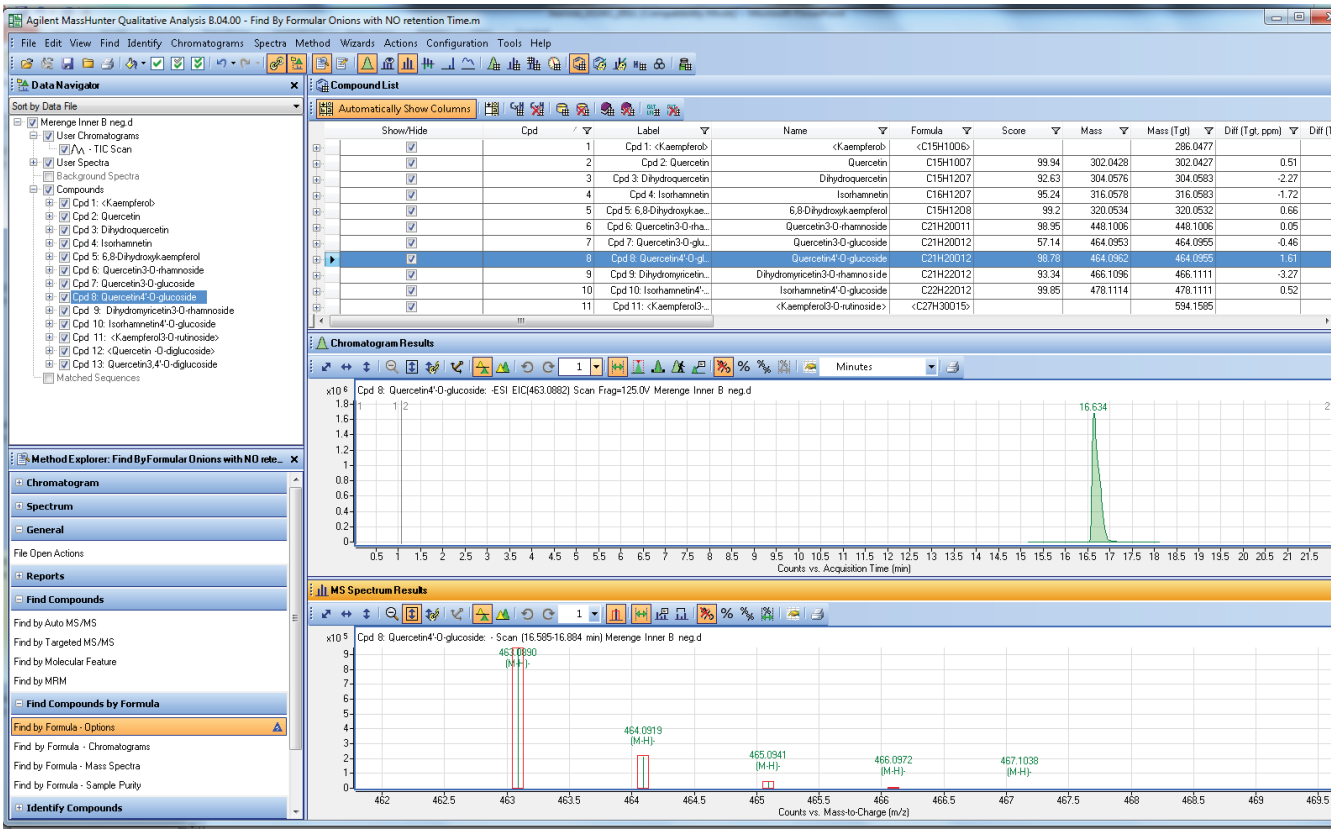


Figure 4. Agilent MassHunter Qualitative Analysis Software was used to compare the measured masses, and isotope abundances and spacing, with those theoretically calculated to generate and score flavonoid matches.

Principal component analysis

Once the list of tentatively identified flavonoids was generated for the different onion varieties, they were further examined for varietal and color difference using the PCA capabilities of MassHunter Mass Profiler Professional Software. PCA was performed in two ways. Both approaches are designed to provide a list of compounds; one tentatively identified from the database search, the other a list of unknowns.

The first PCA performed was based on the matches found in the database. The second approach used molecular feature extraction to extract all possible compounds found in the TIC from all of the varieties. The found unknown compounds were then filtered based on their presence in one variety and absence in another. The molecular feature extractor within MassHunter Qualitative Analysis Software takes all ions that represent chromatographic peaks (thus eliminating background ions) and groups them by adduct clusters, possible isotopes, dimers, trimers, and so forth, all taken together as molecular features without any determination of identity. Each feature is then calculated back to a molecular mass, again without any identification assigned. This unknown approach to mining the data can provide differentiation of varietal differentiation without having to identify the distinguishing compounds.

Results and Discussion

Accurate-Mass Q-TOF LC/MS analysis

Although the TICs from the extracted onion samples appeared somewhat nondescript, they contained many ions representing many compounds. The TIC of the inner layers of the variety Merenge is shown in Figure 5.

Database search results

The MassHunter Qualitative Analysis Software find-by-formula search of the created flavonoid database tentatively identified 19 flavonoids in the seven varieties of onions (Table 3).

Table 3. Flavonoids Tentatively Identified Using the Find-By-Formula Algorithm Provided with Agilent MassHunter Qualitative Analysis Software to Search the Created Database

Delphinidin3-O-(6''-malonyl-glucoside)
Dihydroxykaempferol
Dihydromyricetin3-*O*-rhamnoside
Dihydroquercetin
Isorhamnetin
Isorhamnetin 4'-*O*-glucoside
Kaempferol
Kaempferol 3-*O*-(6"-malonyl-glucoside)
Kaempferol 3,7-*O*-diglucoside
Kaempferol 3-*O*-acetyl-glucoside
Kaempferol 3-*O*-rutinoside
Kaempferol 3-*O*-xylosyl-rutinoside
Quercetin
Quercetin 3,7,4'-triglucoside
Quercetin-*O*-diglucoside
Quercetin 3,4'-*O*-diglucoside
Quercetin 3-*O*-glucoside
Quercetin 3-*O*-rhamnoside
Quercetin 4'-*O*-glucoside

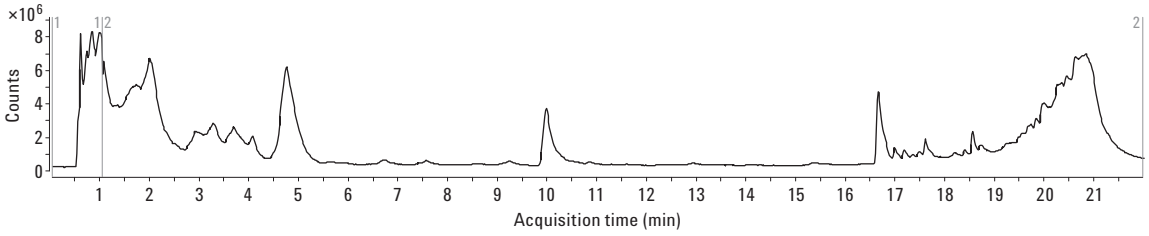


Figure 5. TIC of the variety Merenge.

Principle component analysis results

PCA score plots showed differences in the data from sample to sample. For example, a PCA score plot of the seven varieties of onions, based upon the 19 tentatively identified flavonoids discovered in the database search, showed good separation of the varieties (Figure 6). Because analyses were on the inner layers of the onions, and the extraction solvents excluded anthocyanidins, the distinction of variety was not based on pigment related compounds. The Cowboy and Sommerset varieties group near to each other (less separated), as were Chief, Vaquero, and Salsa. The red varieties, Red Rock and Merenge, were clearly separated from the others.

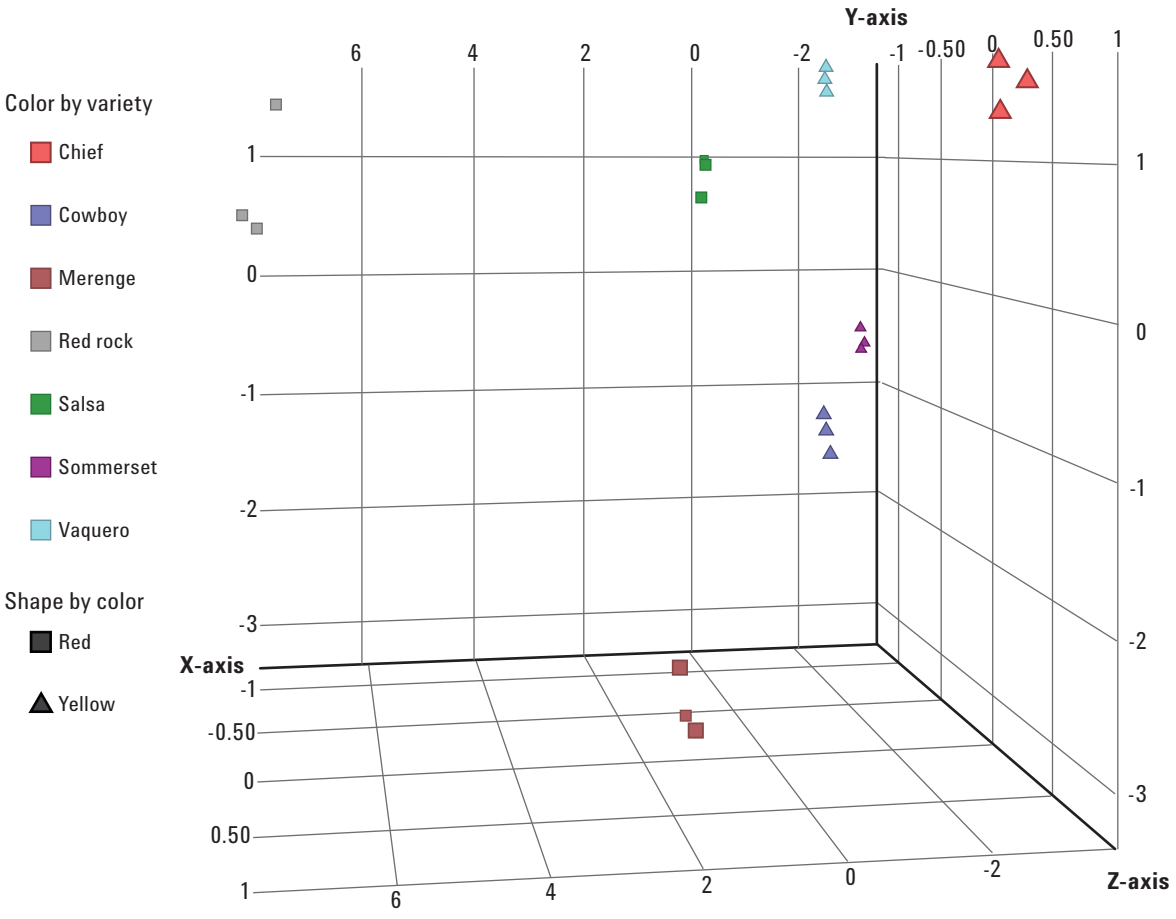


Figure 6. PCA for onion variety difference based on the targeted unknowns found in the flavonoid database search.

The PCA score plot for color difference (Figure 7), showed little distinction between varieties, although there was some correlation along the x-axis.

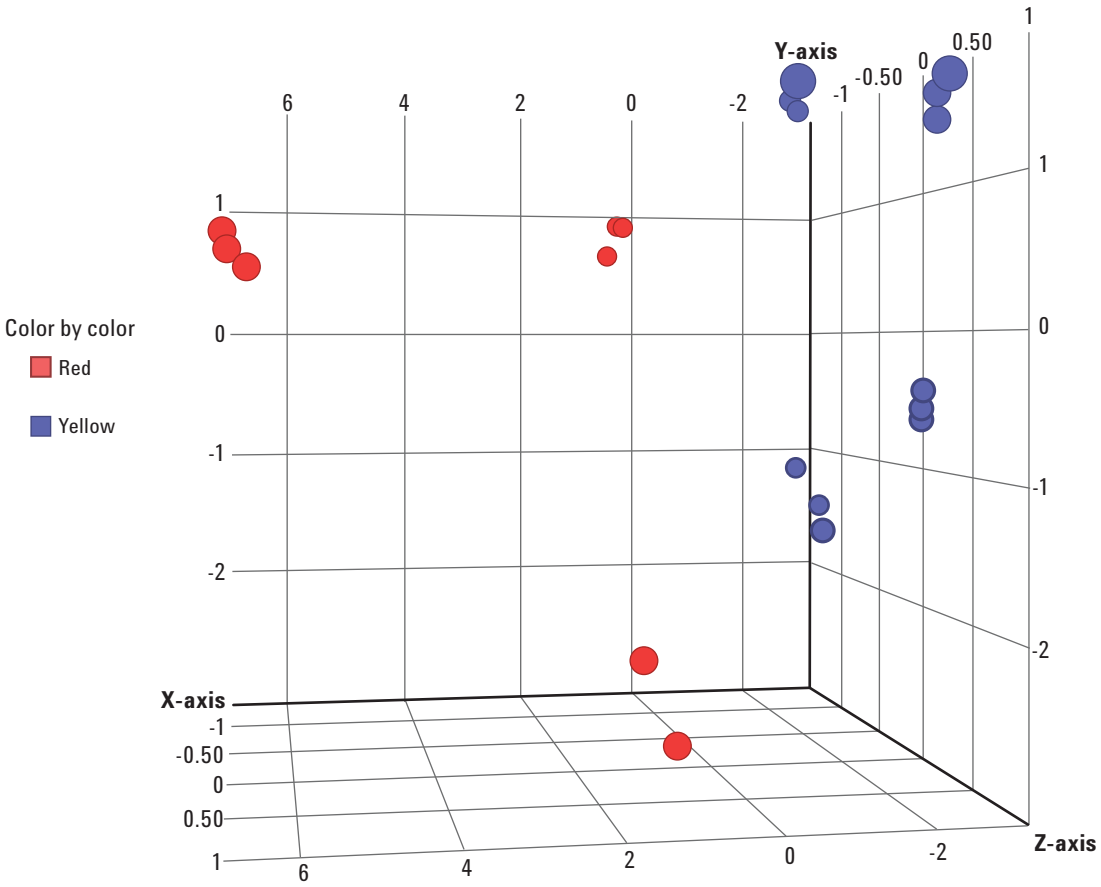


Figure 7. PCA scores plot for color difference based on the tentative compounds found in the flavonoid database search.

A PCA score plot of the seven varieties of onions using all unknown compounds after molecular feature extraction of the data (Figure 8), showed that this approach may also be applicable to evaluating the general composition and color of the different varieties.

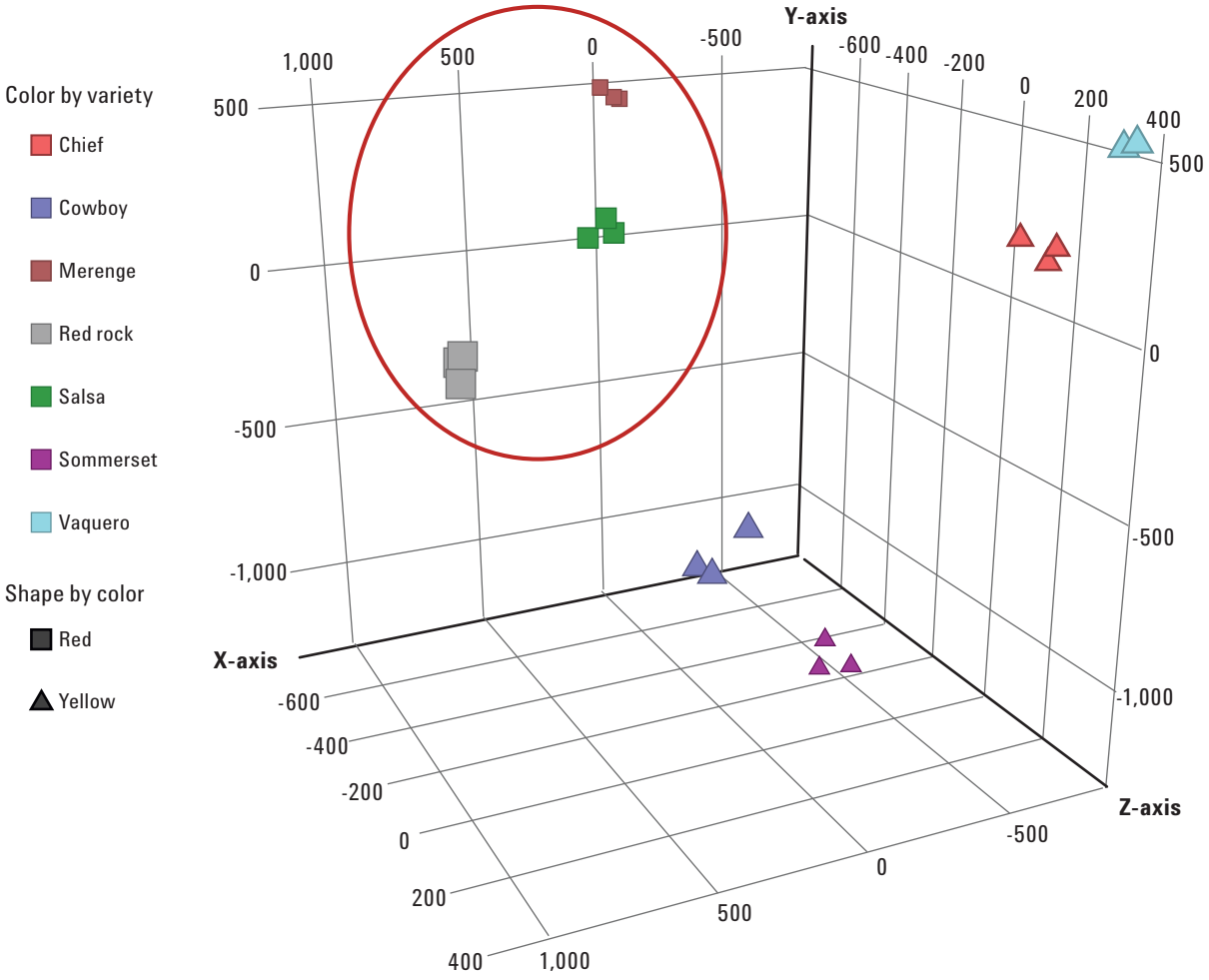


Figure 8. PCA scores plot using all unknown compounds after molecular feature extraction of the data.

The PCA loadings plot shows the entities or compounds responsible for those samples found in the covariance space. The loadings are the list of compounds. Because Chief and Merenge were well separated in the PCA, they could be used to evaluate the loadings and to generate a list of compounds that drive differentiation between the two varieties.

The top 10 compounds tentatively identified as responsible for color differentiation are shown in Table 4. Because there were a number of possible compounds for each formula found by the software, additional MS/MS could be helpful in identifying which compounds are actually present, and that are driving differentiation.

Table 4. Top 10 Nonanthocyanidin Tentatively Identified Compounds Associated with Color Differentiation

6,8-dihydroxy kaempferol
Kaempferol 3-*O*-(6"-malonyl-glucoside)
Kaempferol diglucoside-1
Kaempferol diglucoside-2 (isomer, different retention time)
Kaempferol 3-*O*-acetyl-glucoside quercetin
Quercetin 3,4'-*O*-diglucoside
Quercetin diglucoside-1
Quercetin diglucoside-2 (isomer, different retention time)
Quercetin 3-*O*-rhamnoside
Dihydroquercetin

Conclusion

Using publicly available literature and databases, an exact-mass onion-specific flavonoid database for use in targeted searches was created using Agilent MassHunter PCDL Manager Software. Using this database, a targeted search of the data obtained from Accurate-Mass Q-TOF LC/MS analyses of the extracts of seven varieties of onions identified 19 potential flavonoids. PCA based on the potential flavonoids produced some separation in varietal and color differences. A second PCA using all unknown compounds after molecular feature extraction of the data produced similar results. Using the second PCA approach, the top 10 tentatively identified flavonoids that were more strongly associated with color and variety differences were identified. These results indicate that Accurate-Mass Q-TOF LC/MS, combined with either PCA approach, could be used to establish varietal differences in flavonoid compounds present in onions. However, further analysis of more varieties grown under different conditions is needed to establish clear correlations.

Acknowledgements

The authors would like to thank Gills Onions, LLC, Oxnard, CA for providing onions for this work. We also thank Dr. Susan Ebeler and Dr. Tom Collins of the University of California at Davis and Steven Fischer of Agilent Technologies for invaluable scientific input and help with advanced data processing.

Reference

1. J.S.E. Lee, J.A. Zweigenbaum, A.E. Mitchell "Nontargeted Unknown LC(ESI-)-Q/TOF MS Approaches for Food Verification" *Physical Methods in Food Analysis*, M. Tunick, *et al*. ACS Symposium Series; American Chemical Society. Washington, DC. 2013.

For More Information

These data represent typical results. For more information on our products and services, visit our Web site at www.agilent.com/chem.

www.agilent.com/chem

Published in the USA
November 13, 2013
5991-3419EN

Agilent Technologies