



ELSEVIER

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Profiling of contemporary beer styles using liquid chromatography quadrupole time-of-flight mass spectrometry, multivariate analysis, and machine learning techniques

Hailee E. Anderson ^a, Tiffany Liden ^a, Blair K. Berger ^a, Delphine Zanella ^b, Linh Ho Manh ^c, Shouyi Wang ^c, Kevin A. Schug ^{a, d,*}

^a Department of Chemistry and Biochemistry, The University of Texas at Arlington, 700 Planetarium Place, Arlington, TX, 76019, USA

^b University of Liege, Molecular System, Organic & Biological Analytical Chemistry Group, 11 Allee Du Six Aout, 4000, Liege, Belgium

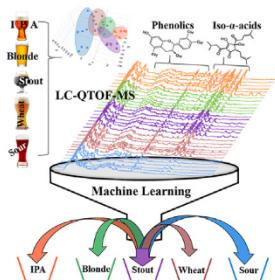
^c Department of Industrial, Manufacturing, and Systems Engineering, The University of Texas at Arlington, 500 West First St., Arlington, TX, 76019, USA

^d Affiliate of Collaborative Laboratories for Environmental Analysis and Remediation, The University of Texas at Arlington, Arlington, TX, 76019, USA

HIGHLIGHTS

- Untargeted determination of chemical components distinguishing beer styles.
- Liquid chromatography – high resolution mass spectrometry and machine learning.
- Analytical and data analysis methods lead to high classification accuracy.
- Unique phenolics and other secondary metabolites identified in different styles.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 15 January 2021

Received in revised form

10 May 2021

Accepted 20 May 2021

Available online 24 May 2021

Keywords:

Craft beer

High resolution mass spectrometry

Principal component analysis

Partial least squares

Discriminant analysis

Untargeted analysis

ABSTRACT

Although all beer is brewed using the same four classes of ingredients, contemporary beer styles show wide variation in flavor and color, suggesting differences in their chemical profiles. A selection of 32 beers covering five styles (India pale ale, blonde, stout, wheat, and sour) were investigated to determine chemical features, which discriminate between popular beer styles. The beers were analyzed in an untargeted fashion using liquid chromatography quadrupole time-of-flight mass spectrometry (LC-QTOF-MS). The separation and detection method were tuned to include compounds from important beer components, namely iso- α -acids and phenolic compounds. Due to the sheer number of unknown compounds in beer, multivariate analysis and machine learning techniques were used to pinpoint some of the compounds most influential in distinguishing beer styles. It was determined that while many phenols and iso- α -acids were present in the beers, they were not the compounds most responsible for the variations between styles. However, it was possible to discriminate each beer style using multivariate analysis. Principal component analysis (PCA) was able to separate and cluster the individual beer samples by style. A combination of statistical tools were used to predict formulas for some of the most influential metabolites from each style. Machine learning models accurately classified patterns in the five beer styles, indicating that they can be precisely distinguished by their nonvolatile chemical profile.

© 2021 Elsevier B.V. All rights reserved.

* Corresponding author. 700 Planetarium Pl., Box 19065, Arlington, TX, 76019-006, USA.

E-mail address: kschug@uta.edu (K.A. Schug).

Abbreviations

principal component analysis (PCA)
partial least squares – discriminant analysis (PLS-DA)
iso- α -acids (IAA)
quadrupole time-of-flight (QTOF)
support vector machine (SVM)
neural networks (NN)
random forests (RF)
Gaussian Naïve Bayes (GNB)

1. Introduction

Despite being one of the oldest and most popular alcoholic beverages across the globe, the identification and classification of beer components still proves to be a challenge due to its high complexity [1,2]. While large industrial brewing companies have focused on standardizing their recipes for mass-production, the contemporary small-batch (so called, “craft”) brewing industry focuses instead on experimental ingredients and procedures to produce even more unique and flavorful beers, which augments the already complex profile [3]. Some beer styles are easily distinguished through physical observations, but the innovation of modern beers by brewers has caused the lines between other styles to blur, making it more difficult to determine into which style a beer fits [4]. For example, American IPA and American Pale Ale, stouts and porters, German Pilsner and Munich Helles, are sometimes so similar in color and flavor that they are difficult to distinguish through purely physical observation. By finding commonalities of beers within the same style on a chemical level, one could more easily determine how to categorize new brews.

Over 75 beer styles exist that can exhibit vast differences in flavor, aroma, and color [5] of this fermented beverage. Despite the wide variations between styles, all beers are brewed with malted grains (typically barley), hops, water, and yeast, thus using the same general procedures and classes of ingredients. Some of the more common styles include India pale ales (IPA), blondes, stouts, wheats, and sours. IPAs are brewed with higher amounts of hops, which contribute a more intense bitter flavor. Blondes are usually light, easy drinking beers with low hop bitterness. Stouts are brewed using roasted grains which results in an incredibly dark color, and usually contain notes of coffee, chocolate, or vanilla flavors. As the name suggests, wheat beers use wheat instead of barley as the primary brewing grain. Sour beers have more of an acidic flavor by utilizing lactic acid-producing bacteria, acetic acid-producing bacteria, or *Brettanomyces* yeasts [5]. These examples show that slight changes in ingredients and procedures can lead to tremendous variations in flavors and fragrances.

One of the ingredients brewers commonly experiment with, both in strain and quantity, are hops, which contribute heavily to the flavor and aroma [6]. Hops introduce a class of compounds known as α -acids. During the wort boil, the α -acids become thermally isomerized to iso- α -acids (IAA), which are largely responsible for the characteristic bitterness of beer, as well as inhibit the growth of Gram-positive bacteria [6]. Another class of chemicals that can affect the flavor of beer are phenolic compounds, which includes polyphenols, phenolic acids, and flavanols, among others [7,8]. Phenolic compounds originate from brewing plant materials, primarily barley and hops [7], but also from many types of fruit that may be used such as cherries, apricots, and oranges [9]. The composition of phenols in beer can vary greatly depending on the ingredients used. In addition to flavor, phenolic compounds

enhance beer stability and exhibit antioxidant activity [10]. Since both the IAA and phenols are major contributors to beer flavor, they have been studied extensively and are known to vary considerably in abundance and type between the beer styles [11–17].

Due to the highly complex nature of beers, targeted analysis with IAA and phenolic compounds alone are not sufficient to be able to differentiate and categorize styles of beer [17]. There are potentially hundreds of other compounds contained in beer that cause differences in styles. Using untargeted analysis and associated techniques would provide the ability to identify some of the key components, which are most influential in differentiating the styles, using multivariate analysis. Principal component analysis (PCA) can be used as a visualization tool to discriminate between beer styles, followed by the application of statistical tools to identify key features.

Moreover, machine learning techniques have gained great success to recognize complex patterns in high dimensional variable space. Different machine learning techniques have been applied to solve challenging data analysis problems in chemistry [18,19]. Specifically, supervised machine learning techniques, such as support vector machine (SVM), neural networks (NN), and random forests (RF), can be used to learn complex data patterns and construct pattern representations to separate and discriminate different chemical data types. In fact, due to its ready availability and complexity, beer has often been used as development and proving ground for various multivariate and chemometric data treatment strategies [20–23].

The objective of this study was to profile different beer styles using liquid chromatography quadrupole time-of-flight mass spectrometry (LC-QTOF-MS) in combination with statistical techniques. These advanced data science techniques could potentially give a more complete picture of the complexity of beer in a way that sensory analysis alone cannot. The combined application of PCA, partial least squares - discriminant analysis (PLS-DA), ANOVA, and various machine learning methods allowed beer styles to be precisely distinguished from one another by identifying the specific compounds responsible for their diversity. This approach could prove valuable for the rapid categorization of new brews, and accentuate the defining characteristics between similar styles, improving our understanding of how ingredients affect the flavor, color, and other properties of this popular beverage. Focusing on providing enhanced ability to distinguish beer styles based solely on their chemical composition can ultimately lead to the development of quality assurance of styles of beer.

2. Materials and methods

2.1. Materials

A single bottle or can of thirty-two U.S. commercial craft beers were purchased from local stores. Five different styles (IPA [I], blonde [Bl], stout [St], wheat [W], and sour [S]) from 22 breweries were represented (Table S1). The beers were freshly opened, and 30 mL were transferred and immediately degassed for 30 min by sonication. Degassed samples were then diluted 50% with water prior to injection. Excess beer was stored at 5 °C for no more than two days.

2.2. Chemicals and reagents

All reagents were of LCMS grade. Water (H_2O), methanol (MeOH), and acetonitrile (ACN) were obtained from Honeywell (Muskegon, MI, USA). Formic acid (98–100%, LCMS grade) was purchased from EMD Millipore (Billerica, MA, USA).

Reference standards including vanillin (99%), 4-hydroxybenzoic acid (99%) (referred to as benzoic acid in the paper), caffeic acid (98%), quercetin (95%), naringin (95%), 4-hydroxycoumarin (98%), myricetin (98%), 4-hydroxy-3-methoxycinnamic acid (99%) (referred to as cinnamic acid in the paper), (+)-catechin (99%), esculin (European Pharmacopoeia reference standard), and chlorogenic acid (European Pharmacopoeia reference standard) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Phenolic standards were chosen based on previous studies involving phenolic compounds in beer [15].

Four variations of the American Society of Brewing Chemists (ASBC) international calibration standards (ICS) for HPLC analysis of isomerized and reduced isomerized α -acids were purchased from Labor Veritas (St. Paul, MN, USA) and are listed in Table 1. The iso- α -acid (IAA), rho-iso- α -acid (RiAA), and hexahydro-iso- α -acid (HiAA) calibration standards were obtained as a purified preparation of the dicyclohexylamine (DCHA) salts. The International Subcommittee for Isomerized Hop α -Acids Standards determined the total percentage of iso- α -acids present in the standards.

2.3. Preparation of standard solutions

Stock solutions of 1 mg mL⁻¹ were prepared for the phenolic and iso- α -acid standards. All stock solutions were prepared in acetonitrile, except for HiAA and RiAA, which were insoluble in ACN and thus, were prepared in methanol. Solutions were stored at 5 °C. Amber vials were used in order to decrease light exposure.

2.4. Preparation of quality controls and samples

In an attempt to limit the influence of instrument variation on sequentially run sample data, special attention was paid to the order of samples analyzed, as well as the inclusion of pooled samples, referred to as QCs [24]. The experimental design, found in Table S2, included 10 quality controls (QC). Five of the QCs were based on beer style, and were prepared by mixing equal parts of each beer from a given style (i.e. IPA-QC, Blonde-QC, Stout-QC, Wheat-QC, and Sour-QC). An “All-QC” was prepared by mixing equal parts of all the beer samples. Lastly, 4 QCs were prepared based on groups. Beers were clustered (one beer from each style) into seven groups [A – G] to ensure that each beer style was not analyzed sequentially. QCs for each group [A-G] were prepared to assist in the visualization of variance within the data set. The group QCs were prepared by mixing equal parts of each beer from the group. All QCs were prepared by diluting the applicable beer mixture by 50% with water. The pooled samples were also used to provide quality assurance that the variation detected was not based on instrumentation, and as a representation of the key features for a given style.

Table 1
Pertinent information about the iso- α -acid standards.

Standard	Abbrev.	Total IAA content (w/w)	trans-	cis-
iso- α -acid (ICS-I4)	IAA	65.2%	t-iCH, t-iH, t-iAH N/A	N/A
rho-iso- α -acid (ICS-R3)	RiAA	65.0%		c-R1iCH, c-R2iCH, c-R1iH, c-R2iH, c-R1iAH, c-R2iAH
hexahydroiso- α -acid (ICS-H2)	HiAA	65.9%	N/A	c-H1iCH, c-H2iCH, c-H1iH, c-H2iH, c-H1iAH, c-H2iAH
tetrahydroiso- α -acid (ICS-T3)	TiAA	99.4%	t-TiCH, t-TiH, t-TiAH	c-TiCH, c-TiH, c-TiAH

2.5. Instrumentation

Analyses were carried out on a Shimadzu Nexera X2 ultra high-performance liquid chromatograph equipped with two solvent delivery pumps [LC-30AD], online degassing unit [DGU-20A5R], autosampler [SIL-30AC], column oven [CTO-20AC], system controller [CBM-20A], and quadrupole time-of-flight (QTOF) mass spectrometer [LCMS-9030] (Shimadzu Scientific Instruments, Inc., Columbia, MD). Separation was achieved using a Restek Raptor C18 column (100 mm × 2.1 mm × 2.7 μ m) (Restek Corporation, Bellefonte, PA). Injection volume was 1 μ L. Mobile phases consisted of water with 0.1% formic acid (solvent A) and acetonitrile with 0.1% formic acid (solvent B). Gradient elution was performed at a flow rate of 0.4 mL min⁻¹, as follows: 0–1 min, 5% B; from 1 to 5 min, 5–75% B; from 5 to 9 min, 75–95% B; and from 9.01 to 15 min, a step gradient back to 5% B for re-equilibration.

Electrospray ionization (ESI) was performed in the negative ionization mode. The MS data was collected under the following ESI conditions: Nitrogen nebulizing gas and drying gas flows were 2 L min⁻¹ and 10 L min⁻¹, respectively; the desolvation line temperature was 250 °C and the heat block temperature was 400 °C; the interface (spray) voltage was –3.5 kV. MRM transitions were optimized for each of the analytical standards, and these settings are detailed in Tables S3 and S4. Samples were also analyzed using data independent acquisition (DIA) in negative ionization mode for untargeted analysis. The untargeted method contained a series of events with the precursors in 15 *m/z* increments ranging from 100 to 1000, with quadrupole 1 (Q1) set with a resolution of 16 to allow for overlap between each precursor. Finally, a collision energy of 25 eV with 17 eV CE spread was used during analysis. The time-of-flight detector was set to scan through various *m/z* ranges shown in Table S5.

2.6. Data processing, statistical analysis, and visualization

LabSolutions was used to acquire the data which were then exported as mzML files and imported into MS Dial (v. 4.00, Yokohama City, Japan). Data processing was performed using MS Dial. The following parameter were used: minimum peak height of 100 counts, mass width of 0.1 Da, mass tolerance of 0.02 Da. A retention time tolerance of 0.1 min and a retention time range of 0–9 min were used as well. The processed data were normalized in MS Dial using the total ion chromatogram (TIC). Results were further exported as a.txt file from MS Dial for further statistical analysis.

Microsoft Excel was used to clean the exported MS Dial results before statistical analysis. Features with a less than a 30% difference from the blank averages were removed from the QCs and samples. Additionally, features with a greater than 30% RSD for each beer sample or QC were removed [24].

After initial data processing and cleaning, data was imported into Metabolanalyst and the areas were normalized using auto-scaling (mean-centered and divided by the standard deviation of each variable). No further data transformation was performed.

MetaboAnalyst 5.0 online (Quebec, Ca) was used to perform statistical analysis. One-way analysis of variance (ANOVA, p -value < 0.05), Random Forest and PLS-DA were used to determine significant features.

MS Finder (v.3.20, Yokohama City, Japan) was used to predict chemical ontologies and formulas for some of the most influential metabolites. An isotopic tolerance of 20% was used in addition to checking the element ratio and the probability of the elements. Elements included in the search were C, H, O, N, P, and S. Database hits for food and natural products were preferentially evaluated [25]. Each experimental MS/MS spectrum was compared to theoretical fragments calculated on known compounds retrieved from structure databases [26].

2.7. Machine learning

The LC-MS data contains high resolution of information including retention time, ion mass spectral data (m/z), and peak areas. To standardize the data structure to train pattern classification models, the LC-MS data was processed and transformed as a 2-dimensional data using a bucketing method similar to the pre-processing step of the PCA analysis. Specifically, for data of each beer, the time domain is from 0 min to 9 min and the m/z domain is from 100 to 1,000. The bucket resolutions for both time domain and m/z domain are 100. For each bucket, the peak areas covered by the bucket was aggregated as the feature of the bucket. In total, there are $100 \times 100 = 10,000$ features extracted to represent each beer data sample. Then, two popular supervised machine learning models, Naive Bayes [27] and random forests [28], were used to learn chromatography feature patterns and discriminate different types of beers. Given the limited number of beer samples, it imposed challenges for the machine learning models to capture discriminative patterns in a high feature dimension of 10,000. Thus, we also performed PCA to transform the raw features into a low-dimensional PCA component space for the pattern classification study.

3. Results and discussion

3.1. Targeted analysis

A targeted analysis of IAA and phenolic compounds was performed for each of the beer samples. A representative standard chromatogram is shown in Fig. S1. Comparison of the MRM traces from the beer samples revealed that a few of the phenolic compounds, such as chlorogenic acid, appeared as multiple peaks in the beer samples, when only one peak was observed in the standard (Figs. S1 and S2). Many of the ingredients used in brewing are biological materials (i.e. yeast, barley, fruits, etc.), which contain complex mixtures of closely related isomers of certain compounds.

Chlorogenic acid (5-CGA) is among these, as multiple isomers such as pseudochlorogenic acid (1-CGA), neochlorogenic acid (3-CGA), and cryptochlorogenic acid (4-CGA) have been identified in various plant materials [29,30]. Therefore, chlorogenic acid and its isomers can be tied back to some brewing ingredients. Fig. S2A represents sour beer #1 (S_1), which was brewed with blackberry puree to add flavor. The predominant hydroxycinnamic acid derivatives found in blackberries are chlorogenic acid (5-CGA) and neochlorogenic acid (3-CGA) [31]. In the chromatogram, the major peak aligns with the chlorogenic acid standard (5-CGA), while the smaller peak matches the major peak in Fig. S2B, which was

hypothesized to be 3-CGA. Fig. S2B represents sour beer #6 (S_6), which is aged on cherries [32]. Sweet cherries are reported to contain neochlorogenic acid as one of the major hydroxycinnamic acids [33]. The major chlorogenic acid peak in the chromatogram is not 5-CGA (the same chlorogenic acid as the standard), so it is likely that it is 3-CGA, the neochlorogenic acid isomer [17].

Principal component analysis (PCA) and hierarchical cluster analysis (HCA) were used to evaluate the potential of using targeted metabolites to distinguish between the beer styles [17]. As expected, Fig. 1B shows there is an increased level of IAA in the IPAs. Diverse levels of IAA can be seen in the blondes, stouts, and wheats, while negligible amounts were detected in the sours. Elevated levels of catechin and its isomer were seen in B_4 and W_6, which caused it to cluster near the IPAs on the heatmap. However, despite the diversity observed in the targeted analysis, it can be seen in Fig. 1 that the beers could not be differentiated using targeted analysis.

3.2. Statistical analysis

Untargeted metabolomics enable the analysis of a wide range of metabolites, and therefore is an ideal tool to highlight key metabolites to differentiate between beer styles. The mass spectral data collected for beer samples was analyzed, and multiple models were assessed to evaluate the influence of instrumental variance.

3.2.1. Samples and quality controls

There were no correlations seen between blanks and quality controls or samples (Fig. S3). A PCA model was used to evaluate the variance before and after removing features with less than a 30% difference from the blanks. Prior to data cleaning, PC 1, 2, and 3 explained 26.7% of the variance. However, after data cleaning PC1 increased from 14.3% to 17% variance. Additionally, the samples are clearly distributed across PC2, with some samples having positive and some having negative correlations to PC2. This suggests that there is variance within the samples that is not observed within the blanks.

Two more unsupervised PCA models were created for the analysis of the beer samples and the QCs (Fig. 2), as well as the beer samples alone (Fig. 3). In Fig. 2A, PC1, PC2, and PC3 represent a total variance of 23%, with each component contributing 11.2%, 7.1%, and 5%, respectively. The points for QC All [yellow], which is a mixture of all the beer samples, are clustered near the zero points for PC1, PC2, and PC3, as would be expected of a composite sample. Additionally, five distinct groupings of the samples and QCs can be seen when marked according to beer style: IPAs [red]; blondes [green]; stouts [purple]; wheats [brown]; and sours [blue]. The highly diverse nature of sour samples is already visible in the initial models that include the QCs and samples (Fig. S3 and Fig. 2). The QCs for 4 beer styles (IPA_QC, BL_QC, St_QC, W_QC) each clustered at the center of a small cloud of the corresponding samples (Figs. S3 and 2). Sours showed the greatest variance within the style compared to the others, which is represented by both the lack of clustering of the samples and the fact that S_QC is located close to All_QC, near the zero point of PC1, PC2, and PC3. This clearly illustrates that the QCs, which are a mixture of the beers from each style, share all the variance within a specific style.

3.2.2. Samples only

After removing the QCs, the same trends are visible (Fig. 3). PC1 explains the greatest amount of variance between the individual beer samples (9.7%), with a total variance of 21.3% for PC1, 2 and 3. IPA is the most unique style of beer, being distinctly separated from all other beer styles. Despite the increased ability to differentiate between beer styles, there are still challenges to separate blondes

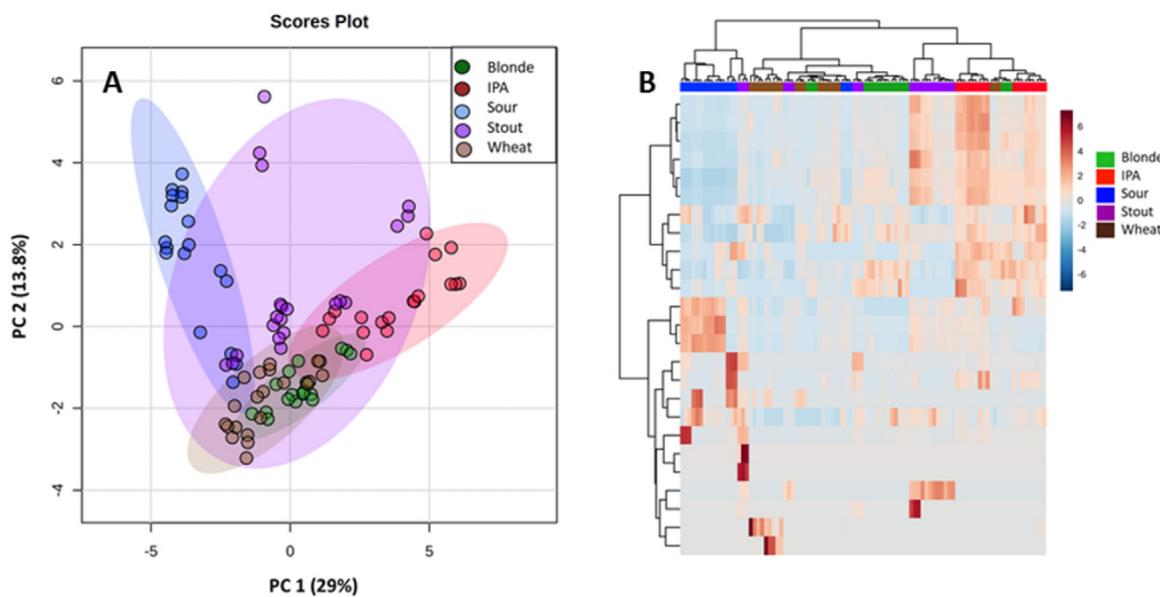


Fig. 1. A Shows a representative PCA and B HCA for the targeted compounds detected using MRM in 32 commercially available beer samples [17].

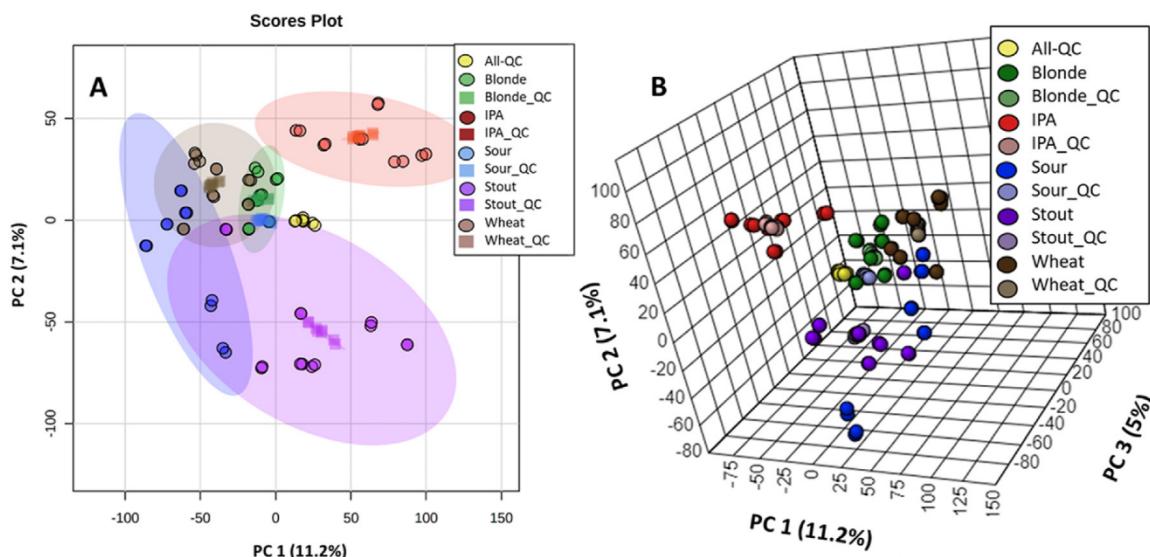


Fig. 2. PCA for 32 commercial beers and pooled quality controls. Points are colored by beer style for visual clarity. [A] PCA 2D model while [B] is a 3D model. Colored by beer style: India pale ale (IPA) [red]; blonde (Bl) [green]; stout (St) [purple]; wheat (W) [brown]; sour (S) [blue]; and quality control of all beer samples (All_QC). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and wheats. Additionally, due to the variation within the styles themselves, outliers can be observed. For example, St_6 is shown to cluster with the blondes in Fig. 3B. Moreover, the heatmap reveals two sours, S_3 and S_5, that are correlated to stouts and IPAs.

3.2.3. Pooled quality controls

Based on the goal of using chemical analysis for style classification quality assurance. Based on the goal of using chemical analysis for style classification quality assurance, pooled samples were used as a representative style sample. Since the QCs are a combination of each sample, the key defining features for each style will thus be accentuated and yield an improved ability to differentiate between beer styles (Fig. S4). Three statistical approaches offered in Metaboanalyst, a univariate (ANOVA) and two

multivariate (PLS-DA and RF), were used in an attempt to remove potential biases from each test and thus reduce the number of key features investigated [34,35]. The top 1,000 features from each statistical approach were further evaluated to include only features shared by all three methods (Fig. 4), in order to make the data manageable. The profile of each peak was investigated to assure gaussian shapes. Such an approach enabled the reduction of the data from 22,200 features to a final dataset of 54 features that have distinctive retention times and accurate masses (Tables 2 and S6).

3.2.4. Tentative identification of features of merit

Although MS/MS spectra were generated for most of the features, the ability to identify them is limited [36]. On average, less than 20% of compounds are identified in most untargeted

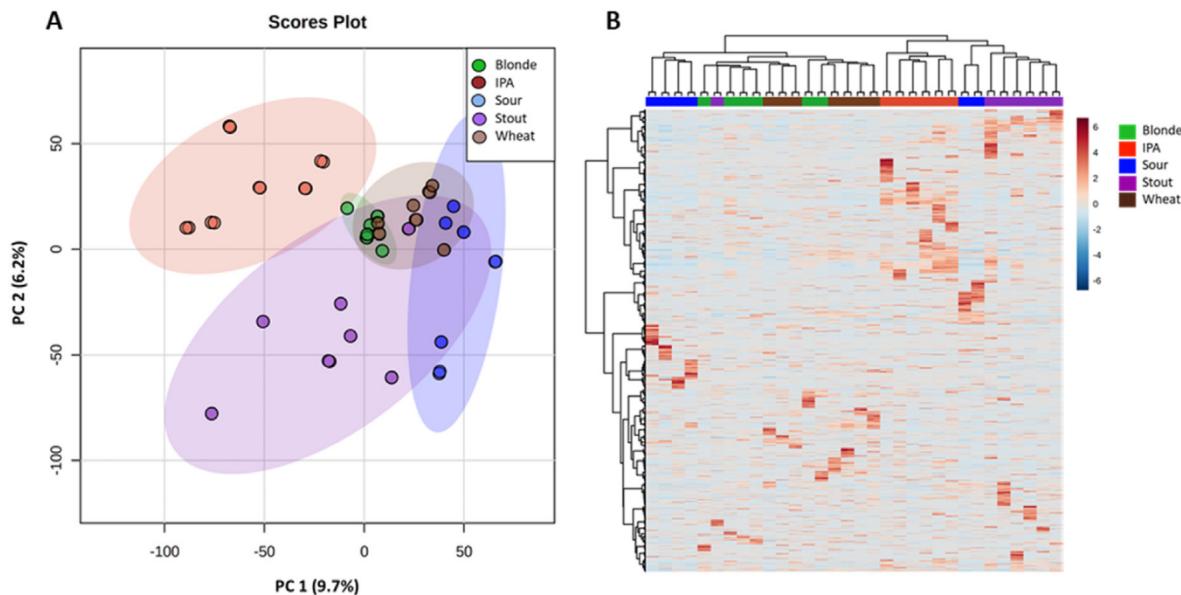


Fig. 3. (A) PCA and (B) HCA for 32 commercial beers. Points are colored by beer style for visual clarity. Colored by beer style: India pale ale (IPA) [red]; blonde (Bl) [green]; stout (St) [purple]; wheat (W) [brown]; and sour (S) [blue]. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

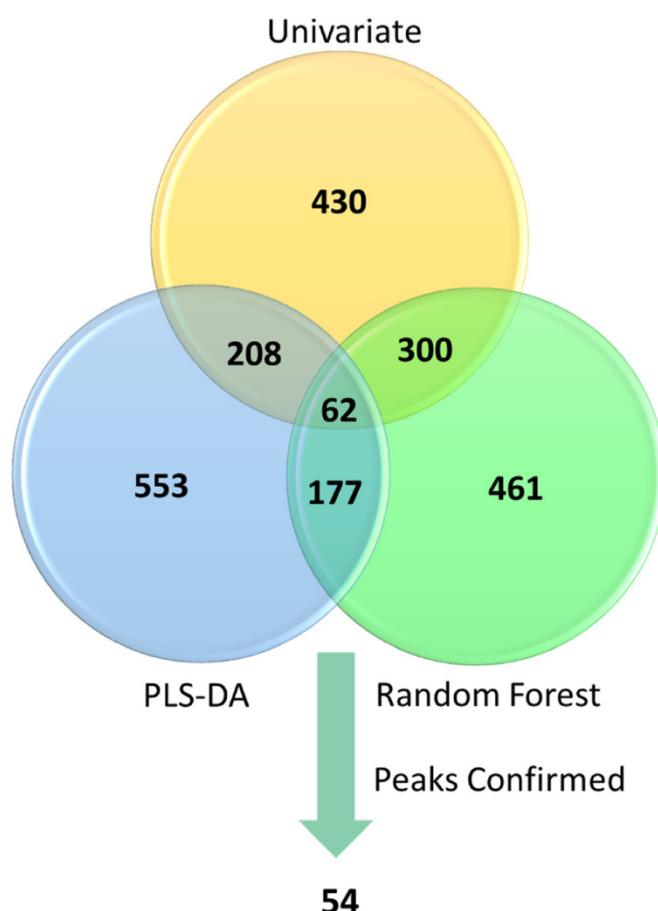


Fig. 4. Workflow for the feature selection, using ANOVA, PLS-DA, and RF, to select influential metabolites able to differentiate between the beer styles.

publications due to limited availability of spectral data in databases [25,26]. Since reference mass spectra were not available, molecular formulas were predicted for features using MS Finder from the precursor ion using the accurate mass, isotope ratio, and product ion information [26]. Theoretical fragmentation was generated for all highlighted features in Table 2, which allowed for predicted molecular formulas and supported the structure elucidation process, thus reaching a level 3 identification based on the Metabolomics Standard Initiative (MSI) [25,37]. However, since the fragmentation of small molecules is still not well understood, theoretical fragmentation was not available to support all predictions [26]. Therefore, the remaining key features, many of which did not have a theoretical fragmentation available for comparison and thus ranking at a level 4, are listed in Table S6.

The key features listed in Table 2 were researched further to determine what their tentative identification could be. Keto acids act as intermediates in the formation of higher alcohols from *Saccharomyces cerevisiae*, a common brewing yeast, during wort fermentation [38]. Based on the mass and structural classification, feature 100 could be 2-oxoadipate, as this has been previously identified in *S. cerevisiae* [39]. Feature 101, a tricarboxylic acid, could possibly be *cis*-2-methylaconitate. This compound was elevated in sours and has been shown to exist in bacteria [40], which are utilized to give sour beers their characteristic flavor. Feature 102, a methoxyphenol found in IPAs, may be methoxyeugenol. Eugenol is known to impart a spicy, clove character in ales [41]. If present in IPAs, it could have possibly been converted to methoxyeugenol during the brewing process. Feature 107 shows that naphthalenes were more prominent in wheat beer, which could be because naphthaleneacetic acid is widely used to increase crop yield and promote growth of wheat and other cereal crops [42–44]. Feature 113 was predicted to belong to the flavone structural class and was found in wheat beers. Various flavones have been identified in wheat and cereal grains [45], which adds confidence to this prediction, but the particular flavone belonging

Table 2

Tentative identification of key feature for beer style quality assurance. Key features were a trifecta match in the top 1,000 most influential features for RF, PLS-DA, and ANOVA. Feature denoted with an * were a trifecta match in the top 500 features. All ions listed are presumed to be [M – H]⁻.

Assigned Number	Style	RT (min)	Accurate Mass	Formula	Error (mDa)	Structural ontology
100	Blonde, IPA	*	0.92	C ₆ H ₈ O ₅	-1.8	Medium-chain keto acids and derivatives
101	Sour (elevated levels)	0.83	187.0277	C ₇ H ₈ O ₆	-2.9	Tricarboxylic acids and derivatives
102	IPA	5.26	193.0906	C ₁₁ H ₁₄ O ₃	-3.5	Methoxyphenols
103	Sour (elevated levels)	2.29	219.0542	C ₁₄ H ₈ N ₂ O	2.2	Indolonaphthyridine alkaloids
104	Wheat	0.65	256.1693	C ₁₇ H ₂₃ NO	1.4	Styrenes
105	Blonde	1.25	304.1088	C ₁₈ H ₁₅ N ₃ O ₂	0.4	Quinazolinamines
106	Blonde, IPA	5.38	317.2608	C ₂₀ H ₃₄ N ₂ O	-1.0	Aminopiperidines
107	Wheat	*	2.83	C ₂₂ H ₃₄ N ₄ O	-4.2	Naphthalenes
108	IPA	8.33	387.2244	C ₁₇ H ₃₂ N ₄ O ₆	0.5	Aminocyclitol glycosides
109	Stout, Wheat	2.74	388.0950	C ₂₁ H ₁₅ N ₃ O ₅	-1.1	Diarylethers
110	Sour, Wheat	*	7.06	C ₂₄ H ₃₈ N ₂ O ₄	4.1	Diterpenoids
111	Wheat	*	2.47	C ₂₉ H ₄₁ NO ₃	-0.5	Steroid esters
112	Stout, Sour, Wheat	*	1.02	C ₃₀ H ₃₀ N ₂ O ₄	-0.5	Pyranquinolines
113	Wheat	0.68	482.1287	C ₂₁ H ₂₅ NO ₁₂	1.7	Flavones
114	Blonde, IPA	*	0.60	C ₂₉ H ₂₉ NO ₁₁	-1.1	Depsides and depsidones

Table 3

The 5-fold cross validation classification performance for the five beer styles.

Methods	extracted features	3 components	5 components	10 components
Naive Bayes	0.70	1.00	0.98	0.98
Random Forest	1.00	0.975	1.00	1.00

to this feature has not yet been identified. Depsides and depsidones, feature 114, were predicted to be key features in IPAs and blondes. This may be due to the presence of gallotannins, a beer stabilizer [46,47], in which the galloyl moieties are linked by depside bonds [48]. Tannins are introduced to beer in the form of hops [49,50], which are used more heavily in the brewing of IPAs. The other key features are still in the process of being identified.

From the untargeted analysis of beer metabolites using the pooled QCs, features allowing the differentiation between styles, based on their presence or absence, were identified. Plots were generated to visualize and compare the influential metabolites between beer styles (Fig. S5). Just as one physical characteristic does not define a beer (i.e. taste, aroma, color, mouthfeel), it would require a combination of key features to chemically identify a style of beer. For example, a combination of the presence of metabolite numbers 100 and 102, in addition to the absence of 103, could be a potential indicator that a beer would qualify as an IPA, as seen in Fig. S5. As expected, due to the easy-drinking and mild flavor characteristics of blondes [5], there were a limited number of metabolites classified as key features in this style. However, the presence of 100, 106, and 114, which are shared features with IPAs, in combination with the presence of metabolite 105, could be developed into a targeted focus for characterizing blondes. Alternatively, stouts, sours, and wheats share some key features. The classification of these could be differentiated from blondes and IPAs by elevated levels of 101 and 103. Furthermore, stouts and sours can be characterized through the presence of metabolite 109 in sours and the absence in stouts. Moreover, the presence of feature 110 in stouts and its absence or reduced levels in sours can provide a distinction between the styles. Finally, wheats have a few additional features, 107 and 113, that would not be expected in stouts and would have lower levels in sours.

3.3. Pattern classification analysis

In order to perform a pattern classification analysis, data from each beer was processed and represented by a feature vector of 10,000 with a resolution of 100 bins in retention time and *m/z*

domains, as described previously. There were 64 samples (two replicates of 32 beers) in total for 5 beer styles. The supervised learning methods used in this study were Gaussian Naïve Bayes (GNB) [27] and Random Forest Classifier (RF) [28] for multiclass pattern classification. For each method, 5-fold cross validation was performed on raw high dimensional feature space and also on the reduced PCA space. The cross-validation accuracies of each method are summarized in Table 3. The GNB model obtained a testing classification accuracy of 70% using the extracted 10,000 features, while it achieved 100% accuracy using the top 3 PCA components. This shows that the GNB model learned discriminative patterns to classify different styles of beers accurately in the low dimensional PCA space. It is notable that the RF model achieved 100% or close to 100% accuracies in both high dimensional feature space and the low dimensional PCA component space. The RF model showed excellent learning capability in a high feature dimension and low sample size setting. In summary, the high classification accuracies indicate that the studied five styles of beer can be characterized by the extracted features and can be precisely distinguished using machine learning models. This provides strong evidence for further investigation on beer styles and their nonvolatile chemical makeup characteristics.

4. Conclusions

Beer is a complex matrix containing hundreds of compounds. Due to the large number of unknown components, targeted analysis alone is insufficient to obtain a full picture of what distinguishes these popular styles, and so untargeted techniques, including PCA and machine learning, were utilized. The untargeted LC-QTOF-MS method and the use of univariate and multivariate statistical analysis was demonstrated to be advantageous for differentiation of beer styles, in addition to highlighting 54 specific metabolites. Molecular formulas and structural characteristics of the metabolites were predicted thanks to the use of high-resolution MS. This combination of untargeted analysis and statistical methods provided evidence that the proof of concept of using chemical analysis as a beer style quality assurance tool is feasible.

The investigation of an expanded beer collection would be

valuable to further develop quality assurance of beer styles. In addition, the chemical validation of the putatively identified metabolites listed in Table 2 would enable to validate their performances to discriminate between the beer styles, and to further develop faster targeted methods to assess beer style quality. Moreover, evaluation of a multifaceted data set using collected results from the analysis of volatiles and nonvolatiles, and further evaluation of the ability to distinguish beer styles using advanced data science techniques, is being pursued.

CRediT authorship contribution statement

Hailee E. Anderson: Conceptualization, Methodology, Validation, Investigation, Writing – original draft. **Tiffany Liden:** Conceptualization, Methodology, Formal analysis, Writing – review & editing. **Blair K. Berger:** Writing – review & editing. **Delphine Zanella:** Methodology, Writing – review & editing. **Linh Ho Manh:** Software. **Shouyi Wang:** Software, Writing – review & editing. **Kevin A. Schug:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Restek Corporation and the Collaborative Laboratories for Environmental Analysis and Remediation (C.L.E.A.R.) at the University of Texas at Arlington for financial support during the course of this research. Thanks is also given to Shimadzu Scientific Instruments, Inc. for providing the instrumentation to perform this research. The Ph.D. fellowship of Delphine Zanella was funded by the Fund for Industry and Agriculture Research (FRIA). We also thank the University foundation Leon Fredericq for the financial support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aca.2021.338668>.

Funding

This work was supported by Restek Corporation.

References

- H.E. Anderson, I.C. Santos, Z.L. Hildenbrand, K.A. Schug, A review of the analytical methods used for beer ingredient and finished product analysis and quality control, *Anal. Chim. Acta* 1085 (2019), <https://doi.org/10.1016/j.aca.2019.07.061>.
- J.U. Fangel, J. Eiken, A. Siersma, H.A. Schols, W.G.T. Willats, J. Harholt, Tracking polysaccharides through the brewing process, *Carbohydr. Polym.* 196 (2018) 465–473, <https://doi.org/10.1016/j.carbpol.2018.05.053>.
- V. Giannetti, M. Boccacci Mariani, P. Torrelli, F. Marini, Flavour component analysis by HS-SPME/GC-MS and chemometric modeling to characterize Pilsner-style Lager craft beers, *Microchem. J.* 149 (2019), <https://doi.org/10.1016/j.microc.2019.103991>.
- M. Naglich, Blurred Lines: Beer Style Competition Categories, Explained, Vinepair, 2019, <https://vinepair.com/articles/beer-style-competition-categories/>. November 7, 2020.
- Beer Styles (n.d.), <https://www.craftbeer.com/beer-styles>. (Accessed 16 October 2020).
- D.D. Keukeleire, *Química nova*, 23(1) (2000) 108, fundam, *Beer Hop Chem* 23 (2000) 108–112.
- M. Lentz, The impact of simple phenolic compounds on beer aroma and flavor, *Fermentation* 4 (2018), <https://doi.org/10.3390/fermentation4010020>.
- T.V. Pai, S.Y. Sawant, A.A. Ghatak, P.A. Chaturvedi, A.M. Gupte, N.S. Desai, Characterization of Indian beers: chemical composition and antioxidant potential, *J. Food Sci. Technol.* 52 (2015) 1414–1423, <https://doi.org/10.1007/s13197-013-1152-2>.
- M. Nardini, I. Garaguso, Characterization of bioactive compounds and antioxidant activity of fruit beers, *Food Chem.* 305 (2020), <https://doi.org/10.1016/j.foodchem.2019.125437>.
- P.M. Aron, T.H. Shellhammer, A discussion of polyphenols in beer physical and flavour stability, *J. Inst. Brew.* 116 (2010) 369–380, <https://doi.org/10.1002/j.2050-0416.2010.tb00788.x>.
- O. Oladokun, A. Tarrega, S. James, K. Smart, J. Hort, D. Cook, The impact of hop bitter acid and polyphenol profiles on the perceived bitterness of beer, *Food Chem.* 205 (2016) 212–220, <https://doi.org/10.1016/j.foodchem.2016.03.023>.
- G. Vanhoenacker, D. De Keukeleire, P. Sandra, Analysis of iso- α -acids and reduced iso- α -acids in beer by direct injection and liquid chromatography with ultraviolet absorbance detection or with mass spectrometry, *J. Chromatogr. A* 1035 (2004) 53–61.
- A. Piazzon, M. Forte, M. Nardini, Characterization of phenolics content and antioxidant activity of different beer types, *J. Agric. Food Chem.* 58 (2010) 10677–10683, <https://doi.org/10.1021/jf101975q>.
- J. Tang, H.A.R. Suleria, F.R. Dunshea, Compounds from Medicinal Plants, Hops and Juniper, 2019, pp. 1–25.
- L. Rehová, V. Škeříková, P. Jandera, Optimisation of gradient HPLC analysis of phenolic compounds and flavonoids in beer using a CoulArray detector, *J. Separ. Sci.* 27 (2004) 1345–1359, <https://doi.org/10.1002/jssc.200401916>.
- E. Parkin, T. Shellhammer, Toward understanding the bitterness of dry-hopped beer, *J. Am. Soc. Brew. Chem.* 75 (2017) 363–368, <https://doi.org/10.1094/ASBCJ-2017-4311-01>.
- H.E. Anderson, T. Liden, B.K. Berger, K.A. Schug, Target profiling of beer styles by their iso-alpha-acid and phenolic content using liquid chromatography - quadrupole - time-of-flight - mass spectrometry, *J. Separ. Sci.* (2021).
- R. Gomez-Bombarelli, A. Aspuru-Guzik, Machine learning and big-data in computational chemistry, in: *Handb. Mater. Model.*, 2020, pp. 1939–1962.
- A.C. Mater, M.L. Coote, Deep learning in chemistry, *J. Chem. Inf. Model.* 59 (2019) 2545–2559, <https://doi.org/10.1021/acs.jcim.9b00266>.
- A. Palmioli, D. Alberici, C. Ciaromelli, C. Airolidi, Metabolomic profiling of beers: combining 1H NMR spectroscopy and chemometric approaches to discriminate craft and industrial products, *Food Chem.* 327 (2020) 1–9.
- L.A. da Silva, D.I. Flumignan, H.R. Pezza, L. Pezza, 1H NMR spectroscopy combined with multivariate data analysis for differentiation of Brazilian lager beer according to brewery, *Eur. Food Res. Technol.* 245 (2019) 2365–2372, <https://doi.org/10.1007/s00217-019-03354-5>.
- L. Mannina, F. Marini, R. Antiochia, S. Cesa, A. Magri, D. Capitani, A.P. Sobolev, Tracing the origin of beer samples by NMR and chemometrics: trappist beers as a case study, *Electrophoresis* 37 (2016) 2710–2719, <https://doi.org/10.1002/elph.201600082>.
- A. Biancolillo, R. Bucci, A.L. Magri, A.D. Magri, F. Marini, Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication, *Anal. Chim. Acta* 820 (2014) 23–31.
- D. Broadhurst, R. Goodacre, S.N. Reinke, J. Kuligowski, I.D. Wilson, M.R. Lewis, W.B. Dunn, Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies, *Metabolomics* 14 (2018) 1–17, <https://doi.org/10.1007/s11306-018-1367-3>.
- I. Blaženović, T. Kind, J. Ji, O. Fiehn, Software tools and approaches for compound identification of LC-MS/MS data in metabolomics, *Metabolites* 8 (2018) 1–23, <https://doi.org/10.3390/metabo8020031>.
- H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihira, W. Tanaka, T. Cajka, K. Saito, O. Fiehn, M. Arita, Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software, *Anal. Chem.* 88 (2016) 7946–7958, <https://doi.org/10.1021/acs.analchem.6b00770>.
- K.P. Murphy, Naive Bayes classifiers, Univ. Br. Columbia, 60 (2006) 1–8, https://doi.org/10.1007/978-3-540-74958-5_35.
- L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- D. Kreml, T. Bajer, P. Bajerova, S. Surmova, K. Ventura, Unremitting problems with chlorogenic acid nomenclature: a review, *Quím. Nova* 39 (2016) 530–533, <https://doi.org/10.5935/0100-4042.20160063>.
- R. Zawirska-Wojtasik, E. Wojtowicz, K. Przygoński, M. Olkowicz, Chlorogenic acid in raw materials for the production of chicory coffee, *J. Sci. Food Agric.* 94 (2014) 2118–2123, <https://doi.org/10.1002/jsfa.6532>.
- L. Kaume, L.R. Howard, L. Devareddy, The blackberry fruit: a review on its composition and chemistry, metabolism and bioavailability, and health benefits, *J. Agric. Food Chem.* 60 (2012) 5716–5727, <https://doi.org/10.1021/jf203318p>.
- Prairie Artisan Ales (n.d.), <https://prairieales.com/beers/>. (Accessed 24 November 2020).
- M.J. Serradilla, M. Fotiric Aksic, G.A. Manganaris, S. Ercisli, D. Gonzalez-Gomez, D. Valero, Cherries: Botany, Production and Uses, CABI, Boston, MA, 2017.
- E. Saccenti, H.C.J. Hoefsloot, A.K. Smilde, J.A. Westerhuis, M.M.W.B. Hendrikx, Reflections on univariate and multivariate analysis of metabolomics data, *Metabolomics* 10 (2014) 361–374, <https://doi.org/10.1007/s11306-013-0598-6>.
- H. Grove, B.M. Jørgensen, F. Jessen, I. Søndergaard, S. Jacobsen, K. Hollung,

- U. Indahl, E.M. Færgestad, Combination of statistical approaches for analysis of 2-DE data gives complementary results, *J. Proteome Res.* 7 (2008) 5119–5124, <https://doi.org/10.1021/pr800424c>.
- [36] J. Guo, T. Huan, Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography-mass spectrometry based untargeted metabolomics, *Anal. Chem.* 92 (2020) 8072–8080, <https://doi.org/10.1021/acs.analchem.9b05135>.
- [37] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.-M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis, *Metabolomics* 3 (2007) 211–221, <https://doi.org/10.1007/s11306-007-0082-2>.
- [38] B.V. Humia, K.S. Santos, A.M. Barbosa, M. Sawata, M. da C. Mendonça, F.F. Padilha, Beer molecules and its sensory and biological properties: a review, *Molecules* 24 (2019), <https://doi.org/10.3390/molecules24081568>.
- [39] L. Palmieri, G. Agrimi, M.J. Runswick, I.M. Fearnley, F. Palmieri, J.E. Walker, Identification in *Saccharomyces cerevisiae* of two isoforms of a novel mitochondrial transporter for 2-oxoadipate and 2-oxoglutarate, *J. Biol. Chem.* 276 (2001) 1916–1922, <https://doi.org/10.1074/jbc.M004332200>.
- [40] cis-2-Methyalaconitate (HMDB0006357), Metabolomics innov. Cent. <https://hmdb.ca/metabolites/HMDB0006357>, 2020. (Accessed 28 April 2021).
- [41] Eugenol Aroxa, Certif. Beer flavour stand. <https://www.aroxa.com/beer/beer-flavour-standard/eugenol/>, 2016. (Accessed 28 April 2021).
- [42] M.A.H.S. Jahan, A. Hossain, J.A.T. Da Silva, A. El Sabagh, M.H. Rashid, C. Barutcular, Effect of naphthaleneacetic acid on root and plant growth and yield of ten irrigated wheat genotypes, *Pakistan J. Bot.* 51 (2019) 1–9, <https://doi.org/10.30848/PJB2019>.
- [43] S.M. Alam, A. Shereen, M.A. Khan, Growth response of wheat cultivars to naphthaleneacetic acid (NAA) and ethrel, *Pakistan J. Bot.* 34 (2002) 135–137.
- [44] B.A. Jeber, H.M. Khaeim, Effect of foliar application of amino acids, organic acids, and naphthalene acetic acid on growth and yield traits of wheat, *Plant Arch.* 19 (2019) 824–826.
- [45] S. Žilić, Phenolic compounds of wheat. Their content, antioxidant capacity and bioaccessibility, *MOJ Food Process. Technol.* 2 (2016) 2–5, <https://doi.org/10.15406/mojfpt.2016.02.00037>.
- [46] G. Belleau, M. Dadić, The Science of Beer Determination of Tannic Acid in Beer by High Performance Liquid Chromatography, 2018, 0470, <https://doi.org/10.1094/ASBCJ-37-0175>.
- [47] M. Dadić, G. Belleau, Determination of tannic acid in beer by thin-layer chromatography, *J. Am. Soc. Brew. Chem.* 36 (1978) 161–167, <https://doi.org/10.1094/asbcj-36-0161>.
- [48] R. Amarowicz, Hydrolysable tannins, encycl, *Food Chem.* 3 (2018) 337–343, <https://doi.org/10.1016/B978-0-08-100596-5.21771-X>.
- [49] P.D. Belur, G. Mugeraya, Microbial production of tannase: state of the art, *Res. J. Microbiol.* 6 (2011) 25–40.
- [50] V.N. Okafor, R.I. Anyalebechi, U.W. Okafor, C.P. Okonkwo, J.N. Obiefuna, M.C. Obiadi, ScienceAssist Phytochemical Constituents of Extracts of Hops and Some Potential Nigerian Hop Substitutes : A Comparative Study in Beer Brewing, 2020, pp. 1–7.