Evaluating the effect of roasting on coffee lipids using a hybrid targeted-untargeted NMR approach in combination with MRI

Kathryn Williamson\textsuperscript{a}, Emmanuel Hatzakis\textsuperscript{b,⁎}

\textsuperscript{a} Department of Food Science and Technology, The Ohio State University, Parker Building, 2015 Fyffe Rd, Columbus, OH, United States
\textsuperscript{b} Foods for Health Discovery Theme, The Ohio State University, United States

ABSTRACT

An integrated targeted-untargeted \textsuperscript{1}H and \textsuperscript{13}C Nuclear Magnetic Resonance (NMR) analysis was applied to determine the impact of roasting on coffee lipids. For targeted analysis, both an internal standard (IS) method, as well as the ERETIC2 tool based on PULCON (Pulse Length-based Concentration determination), were used for quantitation. PULCON allows for quantitative analysis without sample contamination with an IS and was found to be in very good agreement with the traditional IS approach as indicated by a systematic Bland-Altman comparison study. For the untargeted analysis, NMR was coupled with multivariate statistical analysis (MVSA), namely Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), and Orthogonal Projection to Latent Structures Discriminant Analysis (OPLS-DA). \textsuperscript{13}C NMR spectra were acquired using a z-stored spin-echo sequence to achieve higher spectral quality, which is important for both targeted and untargeted analysis. Results showed that roasting has a clear effect on coffee lipids, with diterpenes, oxidation/hydrolysis products and unsaturated fatty acid chains being the most significant markers. In addition, the application of MRI indicated important morphological alterations in bean structure and lipid migration from the endosperm to the surface of the coffee bean.

1. Introduction

Roasting is an integral step in the processing of coffee. The roasting process causes a green coffee bean to undergo physical and chemical changes that result in a porous, brown bean. In addition to the physical changes that a coffee bean undergoes during roasting, its chemical properties also change dramatically. These chemical changes may be the most dominant factor in the development of desirable and complex flavors in a final cup of coffee (Illi & Viani, 2005) and are largely related to Maillard reaction products and carbohydrate degradation (Calligaris, Munari, Arrighetti, & Barba, 2009; Wei et al., 2012). Coffee lipids are also known to have a significant impact on the quality of coffee because of the formation of oxidation products over time (Ortol, Gutierrez, Chiralt, & Fito, 1998). Surprisingly, there are not many studies evaluating the impact of roasting on the lipid fraction of coffee. While it is understood that there is a large influence of roasting on the polar coffee components, some studies indicate that coffee lipids are not significantly affected by the high roasting temperatures (Anesei, De Pilip, Massinif, & Lericp, 2000; Rebeiro et al., 2016). In contrast, other studies have found significant differences in certain lipid components during roasting, although the total lipid profile was not examined (Budryn et al., 2012; Martín, Pablos, González, Valdenebro, & León-Camacho, 2001).

Overall, the literature is conflicting regarding the compositional changes in coffee lipids during roasting. Generating knowledge regarding the impact of roasting on the lipid composition is of specific importance because these compounds and their primary and secondary oxidation products greatly affect the quality and taste/aroma of the commercial value of coffee (Budryn et al., 2012; Toci, Neto, Torres, & Farah, 2013). Further, given the great potential of coffee oil as a value-added material for several industrial and pharmaceutical applications (Raba et al., 2015), it is important to understand the impact of factors such as roasting on its quality and composition.

Here we combined NMR spectroscopy with a hybrid targeted-untargeted multinuclear metabolomic analysis to assess the influence of roasting on the lipid profile of coffee. Targeted analysis was initially performed using an IS. The selection of an appropriate IS can be a difficult task, especially for complex mixtures, as there is an increasing chance for signal overlapping. Additionally, using an IS can contaminate the sample and may prevent further analysis and sample recovery, which is often necessary for biological and clinical studies, as well as sensory analysis studies. An alternative approach is PULCON, in

⁎ Corresponding author at: Department of Food Science and Technology, The Ohio State University, Parker Building, 2015 Fyffe Rd, Columbus, OH, United States.
E-mail addresses: williamson.440@buckeyemail.osu.edu (K. Williamson), chatzakis.1@osu.edu (E. Hatzakis).

https://doi.org/10.1016/j.foodchem.2019.125039
Received 7 May 2019; Received in revised form 16 June 2019; Accepted 17 June 2019
Available online 02 July 2019
0308-8146/ © 2019 Elsevier Ltd. All rights reserved.
which an external standard located into a distinct NMR sample with a predefined concentration is used. This method was first introduced by Wider et al and it is now commercially available as ERETIC2 (Electronic Reference To access In vivo Concentrations 2) tool (Tyburn & Coutant, 2016; Wider & Dreier, 2006). PULCON/ERETIC2 allows for quantification of various compounds without the contamination of the sample with an IS. Although ERETIC2 offers an excellent and convenient tool for targeted mixture analysis, only a few studies utilizing this technique have been conducted so far, mostly in the field of pharmaceutical analysis (Monakhova, Kohl-Himmelseher, Kuballa, & Lachenmeier, 2014; Watanabe et al., 2016) and in food analysis for the determination of polar compounds (Ackermann et al., 2017; Monakhova, Kuballa, Lobell-Behrends, Maixner, & Kohl-Himmelseher, 2013; Monakhova, Lachenmeier, Kuballa, & Mushtakova, 2015). In any case, no systematic statistical study to evaluate its performance has been conducted. In addition, to our knowledge, no application of PULCON/ERETIC2 in lipids has been reported so far.

Most foodomics studies involving NMR are either targeted or untargeted and deal only with $^1$H NMR due to the higher sensitivity of proton nuclei. Our analytical approach allows for simultaneous targeted and untargeted analysis using a single sample preparation that utilizes both $^1$H and $^{13}$C nuclei. The overall goal of this study is to employ $^1$H and $^{13}$C NMR spectroscopy as a non-destructive and reliable tool for the simultaneous targeted-untargeted analysis of coffee lipids before and after roasting. This approach offers an excellent tool for the molecular analysis of lipids in coffee, which can be used for the assessment of coffee oil. The protocol also has the potential to be applied to several other complex matrices. In addition, we present the first MRI application for the analysis of the spatial distribution of lipids in a roasted coffee bean. MRI allows us to study the structural changes in a coffee bean as related to coffee oil sweating and lipid migration due to roasting. To our knowledge, MRI has been used in one previous application to view the water distribution of a green coffee bean (Schmidt, Sun, Litchfield, & Eads, 1996) and this is the first application of MRI for the study of roasted coffee lipids.

2. Materials and methods

2.1. Coffee samples

Eighteen green coffee beans and the corresponding roasted samples, all processed under the same small batch protocol determined by the Specialty Coffee Association of America (SCA) (Specialty Coffee Association, 2015), were used in this study. All samples were harvested in the 2015–2016 harvest year and stored at $-40\, ^\circ$C under the same conditions prior to analysis.

2.2. Chemicals

Chloroform-$d_3$ (CDCl$_3$) and Dimethyl sulfoxide-$d_6$ (DMSO-$d_6$) were purchased from Cambridge Isotope Laboratories (Tewksbury, MA). Chromium acetylacetonate (Cr(acac)$_3$), 3,5-Di-tert-butyl-4-hydroxytoluene (BHT), and chloroform were obtained from Fisher Chemical (Waltham, MA, USA).

2.3. Measurement of water activities and moisture content

Water activities for green coffee beans were measured using a Aqua Lab water activity meter (Meter, Pullman, WA, USA). Moisture content was measured using a vacuum oven at 200 $^\circ$C under vacuum (Isotemp Vacum Oven, Fisher Scientific, Waltham, MA, USA) by weighing the samples before oven drying and after drying for 24 h.

2.4. Sample preparation for NMR experiments

Green and roasted samples were randomized and prepared under an identical protocol as previously described (Williamson & Hatzakis, 2019).

2.5. NMR experiments

$^1$H and $^{13}$C NMR experiments were conducted on a Bruker Avance III spectrometer (Bruker, Ettingen, DE) operating at 700.13 MHz and 176.04 MHz for $^1$H and $^{13}$C nuclei, respectively, equipped with a TXO helium-cooled 5 mm probe. All experiments were performed at 25 $\pm$ 0.1 $^\circ$C and the spectra were processed by the Topspin software package provided by Bruker Biospin.

2.5.1. One-dimensional (1D) NMR spectra

$^1$H NMR spectra were recorded using the following acquisition protocol described previously (Williamson & Hatzakis, 2019). $^{13}$C NMR spectra were obtained with proton decoupling using the inverse gated decoupling version of a z-restored spin echo method (Xia, Moran, Nikonowicz, & Gao, 2008) to minimize NOE effects and produce improved baselines. Repetition delays between pulses were equal to $5 \times T_1$. A spectral width of 221 ppm, 64 K TD, a 90$^\circ$ excitation pulse (9.8 $\mu$s), acquisition time 0.9 s, and relaxation delay of 60 s in order to avoid signal saturation, were used. 64 scans were recorded, and spectra were zero-filled to 64 K. A line broadening of 1 Hz was applied to FIDs before the Fourier Transform.

2.6. PULCON-based quantification

PULCON analysis for the determination of unknown concentrations was performed using a separate sample prepared in the same conditions as the coffee lipid samples that contained a known concentration of BHT. The calculation of unknown concentrations was performed using the ERETIC2 tool from Bruker based on the equation described previously (Monakhova et al., 2014; Wider & Dreier, 2006)

$$C_U = kC_{BHT} \frac{A_U T_U BHT}{A_{BHT} T_{BHT} BHT n_U}$$

where $C$ is the concentration, $k$ is a correction factor dealing with variations in the receiver gain values between the standard BHT sample and the sample for analysis, $A$ is signal area, $T$ is the sample temperature (K), $\theta$ is the pulse length (\(\mu\)s) for a 90$^\circ$ pulse and $n$ is the number of scans. "U" stands for "unknown" and "BHT" stands for "3,5-Di-tert-butyl-4-hydroxytoluene." 0.6 mL of the BHT stock solution was transferred into a 5 mm NMR tube and NMR experiments were performed under quantitative conditions, as described in section 2.6.1. All experiments for PULCON samples and coffee oil samples were processed using the same parameters. Comparisons were made only between samples that were run in the NMR instrument/probe.

2.7. Univariate data analysis

Bland Altman, ANOVA (two way; $F$ ratios and $p$ levels are reported), and paired t-tests (two-tailed, $\alpha$ levels = 0.05), were performed using IBM SPSS Statistics package version 24 (NY, USA).

2.8. Spectral data processing and multivariate data analysis

For $^1$H NMR untargeted analysis, the spectral regions $\delta$ 0.30–12.50 was integrated into regions (bins) with equal width of 0.05 ppm using the AMIX software package (V3.9, Bruker-Biospin). For $^{13}$C NMR based metabolomic analysis, the regions $\delta$ 5–210 and a bin width equal to 0.05 ppm were used. The regions where BHT, water, CDCl$_3$, and DMSO-$d_6$ signals appear were discarded from the data. To adjust for the concentration differences in the lipid fraction among samples, before applying statistical data analysis the bucketed regions were normalized to the total sum of the spectral integrals. MVSA was carried out with SIMCA-P+ software (version 14.1, Umetrics, Sweden). Data were
mean-centered and scaled using Pareto method. Log-transformation was also applied to achieve an improved normal distribution of the data. Principal component analysis (PCA) and orthogonal projection to latent structures with discriminant analysis (OPLS-DA) were conducted on the NMR data. The OPLS-DA model’s confidence level for membership probability was set to 95% and was validated using a 7-fold cross validation method. The quality of the model was assessed by the values of R²Y and Q². Hierarchical cluster analysis (HCA) was applied to determine similarities and/or dissimilarities among coffee lipid samples.

2.9. MRI

All MRI experiments were performed on a Bruker Ascend 750 MHz Micro imaging system (Bruker, Ettlingen, DE). The system had field strength of 17.6 T and was equipped with 89 mm wide bore (WB) and standard bore (SB) systems. A rapid acquisition with relaxation enhancement Rapid Acquisition with Relaxation Enhancement (RARE) sequence was applied to obtain proton density (PD) images with the following parameters: Repetition time (TR) 1 s, echo time (TE) 3.8 ms, slice thickness (SI) 0.50/0.50 mm, field of view (FOV) 1.70/1.20 cm and matrix size 196/128(a).

3. Results and discussion

3.1. Quantitative NMR-validation of PULCON method

3.1.1. Quantitative analysis

1H and 13C NMR have been used previously for the quantification of lipids (Alexandri et al., 2017; Dais, Misjak, & Hatzakis, 2015; Williamson & Hatzakis, 2017) and have been extensively validated by comparison with traditional analytical techniques (D’Amelio, De Angelis, Navarini, Schievano, & Manni, 2013; Knothe & Kenar, 2004; Sacchi, Medina, Aubourg, Addeo, & Paolillo, 1993). For both nuclei, accurate quantification involves the use of a 90° pulse and a delay pulse ≥ 5 × T1 to ensure the complete recovery of the net magnetization. Slow relaxation when utilizing carbon-13 for the analysis is a large issue, as it increases the delay required to ensure full recovery of magnetization between scans. We reduced the T1 relaxation times of the mixture compounds using Cr(acac)3, as a relaxation agent. This rendered the analysis rapid for both 1H and 13C and allowed for the use of an integrated targeted-untargeted approach using both nuclei in one sample preparation. The optimum Cr(acac)3 concentration in order to obtain fast, quantitative results while preserving sufficient spectral resolution for coffee lipid analysis was found to be of 2.4 mM. Under these conditions, the longest T1 relaxation time was 2.4 s for proton and 12 s for carbon-13 nuclei, as measured by an inversion recovery experiment. Increased concentrations of Cr(acac)3 may be used for 13C analysis to further reduce T1 since the signal linewidths in the 13C NMR spectra were found to be surprisingly tolerant to Cr(acac)3 addition. However, they can have a negative impact on the 1H NMR spectrum, which contains important information such as signal multiplicities, which can be lost by the decrease of resolution, demonstrated in Fig. S1. Although similar conditions can be applied for the quantification of lipids from other sources, special care is required when compounds with different chemical structures are present, as these molecules may be characterized by different T1 relaxation times. In addition, the role of the matrix and existence of other paramagnetic metals/cations in the solution can affect T1.

Quantitative 13C NMR analysis requires the elimination of NOE contributions, thus the standard inverse gated decoupling pulse sequence with broadband proton decoupling applied only during the acquisition period is used. However, in modern instruments, especially ones that operate in Larmor frequencies higher than 600 MHz and are equipped with cryoprobes, 13C NMR analysis suffers from baseline and phasing issues, which can cause problems in targeted and untargeted analysis. For that reason, we performed 13C NMR experiments using a z-restored spin-echo pulse sequence, which produces spectra with smoother baselines compared to standard 13C (1H) experiments (Xia et al., 2008), as shown in Fig. S2. The z-restored spin-echo sequence minimizes first-order phase errors and baseline humps that occur as a result of the high Q-factor of modern NMR probes, especially cryoprobes. In addition, the use of adiabatic pulses included in the sequence allows for uniform excitation of all frequencies, even in cases of wide bandwidths which are becoming more common as high field instruments are more readily available. Offset effects can cause significant errors in 13C NMR analysis and create limitations in signal integrations, as only signals that appear in similar spectral regions can be compared (D’Amelio et al., 2013). The use of adiabatic pulses (Crp80,0,5,20,1 and Crp80comp,4) for inversion and refocusing, ensures the elimination of off-resonance issues and the equal excitation of all frequencies. Here, we used an inverse gated decoupling version of the z-restored sequence. To our knowledge, this is the first application of this experiment in lipid analysis and quantitative analysis in general.

Table S1 summarizes the compounds that can be determined quantitatively, in a non-destructive manner, by 1H and 13C NMR using specific diagnostic signals with BHT as an IS. The most important components that can be quantified by NMR are oleic acid (OL), linoleic acid (LO), linolenic acid, saturated fatty acids (SFA), kahweol (K), cafestol (C), caffeine (Cf), diglycerides (DG), free fatty acids (FFA) and various sterols. 1H and 13C NMR are in a good agreement, as found by the comparison between 1H and 13C NMR for the calculation of various compounds such as K, Cf and LO (Table S2), confirming the quantitative nature of the z-stored experiment. A systematic deviation was found between 1H and 13C NMR for the calculation of LN, which can be attributed to the low S/N in the carbon spectrum, unrelated to the z-store sequence. Using this method, the targeted simultaneous analysis of certain compounds (Table S1) in coffee lipids can be achieved in less than 3 min with 1H NMR and in about 1 h with 13C NMR, while at least one day may be required when using traditional methods. The two analyses are complementary to each other: 1H offers a fast quantification of many lipid compounds, while 13C provides spectra with higher resolution.

3.1.2. PULCON validation

Next, we performed a systematic comparative study to assess the agreement between PULCON and the regular IS approach using the Bland-Altman methodology. To our knowledge, this is the first application of PULCON/ERETIC2 for lipid quantification, as well as the first attempt to systematically investigate its efficiency in small molecules using a robust statistical approach. The two most critical experimental parameters that must be considered for minimizing quantification errors are the pulse length and the receiver gain value. Quantification with PULCON seems to be sensitive to pulse miscalibration, and even small variations (eg 1 us) in the pulse length can cause significant deviation in the calculated concentration of the analyte. This is not surprising, as the concentration being determined has a direct dependence on the pulse length used for the acquisition of the reference and the unknown spectra as shown in Eq. (1). For this reason, the NMR probe should be tuned and matched for all samples, because the efficiency of the delivery of the RF power to the sample for a given pulse length value depends on tuning and matching.

The signal area in an NMR experiment is also dependent on the receiver amplifier gain (RG) value, a parameter that matches the magnitude of the time domain NMR signal (FID) to the dynamic range of the analogue to digital converter (ADC) (Frank, Kreissl, Daschner, & Hofmann, 2014). Our results indicate that the concentration calculated with 13C NMR and PULCON for a given sample has a small dependence on the RG value (Fig. S3A). This is in contrast to a previous study (Frank et al., 2014), which didn’t find a clear correlation between RG value and concentration by PULCON. However, in that study, samples containing pure compounds were used and 1H nuclei were tested. To examine if this relationship is due to PULCON, we also examined the
relationship between the concentration calculated using IS and the RG value (Fig. S3B). Results indicated that there is not a specific pattern, thus the PULCON method is more susceptible to RG variations compared to IS. The S/N was also found to have a linear relationship with RG values, as expected (Fig. S3C).

To validate the reliability of the PULCON method, we quantified several components of the lipid fraction of green coffee, namely LO, LN, SFA, K, C, DGs, FFA and Sterols, using the traditional IS method as well as PULCON. Representative quantitative data for $^1$H NMR and $^{13}$C NMR analysis are shown in Table 1. Based on the data in these tables, we generated linear regression equations of the type:

$$(\text{Analyte, PULCON}) = a + b(\text{Analyte, IS})$$

to evaluate the correlation between the data obtained with PULCON and those obtained with an IS. Table S3 displays the linear regression data for the dependent variable (PULCON) with the independent variable (IS) for the analysis of LO, Campesterol and C. All of the linear regression models had $R^2$ values above 0.9, showing that PULCON and the IS method are well correlated. Because linear regression only correlates the two methods and does not assess the existence of any agreement between them, we performed a Bland-Altman comparison for the main compounds that appear in coffee oil. Bland-Altman is an alternative approach to linear regression that is used for evaluating the agreement between them, we performed a Bland-Altman comparison and other non-polar compounds, although some literature regarding these changes is conflicting (González, Pablos, Martín, León-Camacho, & Valdenebro, 2001; Martín et al., 2001; Raba et al., 2015; Rebeiro et al., 2016). It is expected that these changes will result in the disappearance of certain functional groups and the appearance of new ones. Further, it is expected that these changes are detectable via NMR spectroscopy. Targeted (IS or PULCON) and untargeted NMR analysis can be applied to investigate the impact of roasting on coffee lipids. Upon looking at the green and roasted coffee bean lipid spectra, there appear to be many spectral similarities, but a closer inspection of the spectra reveals remarkable differences. For example, Fig. 2 shows the expanded spectral region of the $^1$H NMR spectrum from $\delta$ 12.5–8 for the coffee lipids of a green bean and the lipids extracted from the same sample after roasting. It can be visually observed that roasting has an impact on the oxidation product profile, as the formation of aldehydes upon the degradation of hydroperoxides and other unsaturated chains occurs. These aldehydes have been associated with the characteristic aroma of roasted coffee (Calligaris et al., 2009).

### 3.2. Effect of roasting on coffee lipids composition

#### 3.2.1. Univariate analysis

Roasting is likely to cause compositional changes in coffee lipids and other non-polar compounds, although some literature regarding these changes is conflicting (González, Pablos, Martín, León-Camacho, & Valdenebro, 2001; Martín et al., 2001; Raba et al., 2015; Rebeiro et al., 2016). It is expected that these changes will result in the disappearance of certain functional groups and the appearance of new ones. Further, it is expected that these changes are detectable via NMR spectroscopy. Targeted (IS or PULCON) and untargeted NMR analysis can be applied to investigate the impact of roasting on coffee lipids. Upon looking at the green and roasted coffee bean lipid spectra, there appear to be many spectral similarities, but a closer inspection of the spectra reveals remarkable differences. For example, Fig. 2 shows the expanded spectral region of the $^1$H NMR spectrum from $\delta$ 12.5–8 for the coffee lipids of a green bean and the lipids extracted from the same sample after roasting. It can be visually observed that roasting has an impact on the oxidation product profile, as the formation of aldehydes upon the degradation of hydroperoxides and other unsaturated chains occurs. These aldehydes have been associated with the characteristic aroma of roasted coffee (Calligaris et al., 2009).

When comparing the signal integrals obtained by the entire aldehyde region, $\delta$ 10.5–9.0, there was a difference between green and roasted samples (Fig. 2), however, it was not statistically significant ($F_{1,24} = 3.365, p = 0.079, Table 2$) as found by 2-way ANOVA. The effect of geographical origin on the lipid components was not the objective of this study. However, coffee samples came from 11 different growing regions, therefore region is a factor that cannot be ignored. Thus, a two-way ANOVA was conducted to assess geographical origin’s impact on the samples used in this study. No significant difference was the

<table>
<thead>
<tr>
<th>Sample</th>
<th>LO ($^1$H NMR)</th>
<th>LN ($^1$H NMR)</th>
<th>Campesterol ($^1$H NMR)</th>
<th>Cafestol ($^1$H NMR)</th>
<th>Kahweol ($^{13}$C NMR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BHT PULCON</td>
<td>BHT PULCON</td>
<td>BHT PULCON</td>
<td>BHT PULCON</td>
<td>BHT PULCON</td>
</tr>
<tr>
<td>1</td>
<td>62.25 62.41</td>
<td>2.60 2.53</td>
<td>0.43 0.43</td>
<td>5.90 5.96</td>
<td>11.93 9.96</td>
</tr>
<tr>
<td>2</td>
<td>96.20 97.80</td>
<td>3.91 3.95</td>
<td>0.64 0.65</td>
<td>8.76 8.89</td>
<td>11.82 12.04</td>
</tr>
<tr>
<td>3</td>
<td>126.50 126.91</td>
<td>5.35 5.34</td>
<td>0.68 0.65</td>
<td>11.90 11.90</td>
<td>12.65 12.24</td>
</tr>
<tr>
<td>4</td>
<td>77.54 83.40</td>
<td>3.46 3.93</td>
<td>0.43 0.47</td>
<td>9.78 9.14</td>
<td>9.68 9.23</td>
</tr>
<tr>
<td>5</td>
<td>109.92 105.88</td>
<td>3.79 3.63</td>
<td>0.73 0.70</td>
<td>10.94 10.57</td>
<td>11.30 11.2</td>
</tr>
<tr>
<td>6</td>
<td>60.97 63.36</td>
<td>2.68 2.73</td>
<td>0.55 0.58</td>
<td>6.44 6.75</td>
<td>15.27 13.2</td>
</tr>
<tr>
<td>7</td>
<td>90.81 91.89</td>
<td>3.64 3.74</td>
<td>0.77 0.73</td>
<td>11.36 11.51</td>
<td>11.78 10.09</td>
</tr>
<tr>
<td>8</td>
<td>92.93 89.94</td>
<td>3.55 3.44</td>
<td>0.75 0.73</td>
<td>11.32 10.96</td>
<td>9.25 9.30</td>
</tr>
<tr>
<td>9</td>
<td>86.06 84.93</td>
<td>3.34 3.19</td>
<td>0.51 0.50</td>
<td>9.80 9.65</td>
<td>7.23 6.80</td>
</tr>
<tr>
<td>10</td>
<td>99.16 99.55</td>
<td>3.71 3.73</td>
<td>0.81 0.80</td>
<td>6.96 6.99</td>
<td>14.62 12.38</td>
</tr>
<tr>
<td>11</td>
<td>74.84 76.15</td>
<td>2.89 2.94</td>
<td>0.33 0.33</td>
<td>6.98 7.10</td>
<td>5.02 5.56</td>
</tr>
<tr>
<td>12</td>
<td>82.01 80.32</td>
<td>3.05 2.99</td>
<td>0.44 0.43</td>
<td>6.60 6.46</td>
<td>8.39 8.40</td>
</tr>
<tr>
<td>13</td>
<td>97.84 95.69</td>
<td>3.13 3.42</td>
<td>0.57 0.56</td>
<td>10.46 10.24</td>
<td>11.52 10.21</td>
</tr>
<tr>
<td>14</td>
<td>74.84 75.33</td>
<td>3.06 3.07</td>
<td>0.63 0.62</td>
<td>12.34 12.43</td>
<td>7.81 6.74</td>
</tr>
<tr>
<td>15</td>
<td>62.28 67.82</td>
<td>3.13 3.11</td>
<td>0.48 0.48</td>
<td>8.20 8.14</td>
<td>6.55 7.46</td>
</tr>
<tr>
<td>16</td>
<td>73.69 71.02</td>
<td>3.06 2.84</td>
<td>0.41 0.40</td>
<td>9.70 9.32</td>
<td>7.81 7.18</td>
</tr>
<tr>
<td>17</td>
<td>72.94 72.05</td>
<td>2.90 2.86</td>
<td>0.55 0.55</td>
<td>9.40 9.29</td>
<td>8.02 7.81</td>
</tr>
<tr>
<td>18</td>
<td>74.05 75.17</td>
<td>3.37 3.42</td>
<td>0.45 0.46</td>
<td>8.36 8.50</td>
<td>6.11 6.60</td>
</tr>
<tr>
<td>Mean</td>
<td>84.16 84.42</td>
<td>3.38 3.38</td>
<td>0.56 0.56</td>
<td>9.18 9.10</td>
<td>9.82 9.24</td>
</tr>
<tr>
<td>Std.</td>
<td>16.99 16.12</td>
<td>0.61 0.64</td>
<td>0.13 0.13</td>
<td>1.96 1.89</td>
<td>2.85 2.31</td>
</tr>
</tbody>
</table>
observed among coffee bean geographical origins when analyzing the total aldehyde region ($F_{1,24} = 0.407, p = 0.930$). Despite the fact that no statistically significant differences were found between green and roasted samples when analyzing the aldehyde region, statistically significant variations were observed when only the aldehyde doublet at $\delta 10.12$ was examined ($F_{10,24} = 6.336, p = 0.019$), confirming that there are differences in oxidation products of green and roasted samples. The significant increase in this aldehyde after roasting indicates that oxidation of both FA and primary oxidation products does occur to some degree during roasting, leading to the production of secondary oxidation products. When conducting a two-way ANOVA with the aldehyde doublet to examine the influence of geographical origin, Vietnam coffee was found to have significantly greater levels of the $\alpha,\beta$-unsaturated aldehyde ($F_{10,24} = 2.827, p = 0.018$), compared to Kenya, Colombia, and Rwanda ($p = 0.033, p = 0.025, p = 0.020$, respectively). However, studies using a large number of samples that differ only by geographical area are necessary before this relationship can be confirmed and fully understood.

In addition to FA-induced oxidation products, roasting also has a significant impact on the coffee diterpenes K and C. There are contradictory results in literature regarding the effect of roasting on these coffee-specific terpenes. They have been found to decrease during roasting by some studies (Kölling-Speer, Strohschneider, & Speer, 1999), while others have reported that they remain stable (Urgert et al., 1995). In our study, both terpenes were found at significantly lower concentrations in roasted coffee compared to green coffee, indicating degradation during roasting. When conducting a two-way ANOVA, K was found to significantly decrease after roasting ($F_{1,24} = 14.138, p < 0.001$), as measured by the signal of H19 at $\delta 7.272$. The growing region also had a significant impact on K’s concentration in some instances ($F_{10,24} = 16.201, p < 0.001$). Differences among these samples are presented in supplemental Table S4. A similar pattern was observed when measuring C among the coffee samples. Roasting caused a significant decrease in C, as measured by the H19 at $\delta 7.22$ ($F_{1,24} = 8.043, p = 0.009$). Additionally, geographical origin of the coffee beans had a significant effect on C concentration in some of the samples ($F_{10,24} = 4.471, p = 0.001$). There was not an observed pattern among these geographical differences, however this finding may be important for future authentication studies aiming at identifying coffee based on geographical origin (Table S4).

The decreased diterpene concentrations found when the aromatic

---

Table 2

<table>
<thead>
<tr>
<th>Spectral Area</th>
<th>Sample</th>
<th>Mean Peak Area ($\times 10^3$)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha,\beta$-unsaturated dienal ($\delta 10.12$)</td>
<td>Green</td>
<td>0.281 ± 0.170</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>Roasted</td>
<td>0.398 ± 0.70</td>
<td></td>
</tr>
<tr>
<td>Total aldehydes ($\delta 10.5-9.0$)</td>
<td>Green</td>
<td>1.15 ± 0.530</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>Roasted</td>
<td>1.69 ± 0.986</td>
<td></td>
</tr>
<tr>
<td>K,C ($\delta 4.24-4.19$)</td>
<td>Green</td>
<td>25.1 ± 2.57</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Roasted</td>
<td>23.05 ± 2.14</td>
<td></td>
</tr>
<tr>
<td>FFA ($\delta 175.5$)</td>
<td>Green</td>
<td>2.17 ± 0.730</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Roasted</td>
<td>2.90 ± 0.865</td>
<td></td>
</tr>
<tr>
<td>LN ($\delta 2.81-2.78$)</td>
<td>Green</td>
<td>0.830 ± 0.0722</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Roasted</td>
<td>0.735 ± 0.0680</td>
<td></td>
</tr>
</tbody>
</table>

---

Fig. 1. Difference (bias) plots obtained from Bland Altman analysis of green coffee oil samples analyzed by IS and PULCON methods for (A) LO, (B) Campesterol, and (C) Cafestol.

Fig. 2. The effect of roasting on the oxidation status of coffee oil. Comparisons between coffee oil extracted from green beans (A) and roasted beans (B).

![Fig. 1](image1.png)

![Fig. 2](image2.png)
hydrogen is used for measurement is likely due to the susceptibility of their double bonds to oxidation during roasting. Similar results are found when using the integrals of the diastereotopic protons H17a and H17b of K and C having signals at δ 4.428 and δ 4.212, respectively. These protons are located close to the ester group of esterified K and C, indicating that these two terpenes may also be subjected to hydrolysis during roasting. The formation of Dehydrokahweol and Dehycrofasteol during roasting has been previously reported (Kolling-Speer, Kurt, Thu, & Speer, 1997), however their presence was not detected in this study as the characteristic signal of H15 at δ 5.45 (Scharnhop & Winterhalter, 2009) was not observed in the spectrum.

Roasting also seems to accelerate the hydrolysis of triglycerides in coffee oil. When comparing the quantity of FFA between green and roasted coffee, a significant increase in FFA following roasting was observed (F(1,24) = 9.683, p = 0.005, Table 2), which is in agreement with previous studies (Raba et al., 2015; Speer & Kölling-Speer, 2006). Additionally, a two-way ANOVA found no significant differences in FFA among samples based on region (F(10,24) = 2.213, p = 0.373) with previous studies (Raba et al., 2015; Speer & Kölling-Speer, 2006).

Another finding, 1,3-DG during roasting, or forming the signal (Q2 < 0.3).

A comparative OPLS-DA analysis was also performed for the two groups, green and roasted, to focus on the variance originating from the effect of roasting. The cross-validated scores plots showed a clear separation between them, indicating significant differences in the global biochemical compositions between the two groups. R2Y and Q2, the most important quality indicators for OPLS-DA (Fig. 3C), were 0.98 and 0.92, respectively, indicating very high goodness of fit and model predictability, and reliable distinction obtained by the model. The model was further validated using CV-ANOVA (Table S5) and permutation tests (Fig. S4) to ensure that there is no overfitting (permuted model R2 = 0.70, Q2 = −1.64). Although OPLS-based models are considered reliable and established chemometric tools for supervised analysis, they remove the non-predictive variance. Thus, a PLSDA model that lacks the orthogonal signal correction filter was also used. The PLS-DA model also had very good quality indicators (R2 = 0.94, Q2 = 0.80), further confirming the significant impact of roasting on coffee lipids. The S-plot that displays the modeled covariance (p[1]) versus the modeled correlation (p[corr]) of metabolites (bins) is shown in Fig. 3D. Bins that have an important impact on group classification are terpenes (e.g. Kahweol, bin at δ 6.275), oxidation products (e.g. aldehyde, bin at δ 9.725) and unsaturated fatty acid chains (e.g. LO, bin at δ 2.275), which is in agreement to the findings from univariate analysis. There are also important bins contributing to this separation that belong to minor compounds not identified in this study.

MVSA using the 13C NMR data also showed good separation between green and roasted samples on the PCA model (Fig. S5), although one strong outlier was detected. MVSA also showed a reliable classification in the OPLS-DA model (R2 = 0.97, Q2 = 0.87). Although 13C NMR provides higher spectral resolution compared to 1H, we found that this does not necessarily increase the predictability of the multivariate model. This is because some compounds such as oxidation products, that appear in minor concentrations, but are good predictors in the 1H spectrum, they appear in coffee oil in concentrations lower than the detection limit of 13C NMR, or only appear as weak signals in the 13C spectrum. The signals of minor compounds are further suppressed due to the scaling to the total intensity, which is performed prior to the MVSA and is necessary for correcting weight-related variations. The most significant descriptors identified by 13C NMR are unsaturated fatty acyl chains. This finding is in agreement with previous studies utilizing gas chromatography (Martín et al., 2001). 13C NMR-based analysis was also able to identify terpenes as descriptors.

The dependence of the green and roasted coffee bean lipid metabolome on aω was also evaluated using MVSA, but no significant correlation was observed. However, further studies using monovarietal samples of a single geographical origin in order to eliminate additional factors of variance are required to confirm these results. Although moisture content has been found to have an effect on coffee lipids (Ortol et al., 1998), it seems that it has small or no influence during roasting. The effect of aω on the composition of oxidation and hydrolysis products in green and roasted coffee beans was examined, however no correlation was observed for the samples used in this study, as low values for Q2 were observed for these models.

3.3. Physical transformation in the coffee bean due to roasting

Both targeted and untargeted analysis revealed significant differences in the composition and structure of several major coffee lipids and other minor compounds during roasting. Roasting also affects the structure of coffee bean and changes in oil distribution. To observe these structural changes, MRI was employed. A rapid acquisition with relaxation enhancement RARE sequence (Henning, Nauerth, & Friedburg, 1986) was used to acquire proton density (PD) images of green and roasted Arabica coffee beans. PD MRI with the same
acquisition parameters for green and roasted beans allowed for the comparison of PD distribution in green (Fig. 4A) and roasted (dry) (Fig. 4B) coffee beans. In Fig. 4, the mobility of protons is visible in grayscale with the brighter pixels indicating the presence of highly mobile protons, and the darker areas representing either a decrease in mobility or absence of protons (void). It appears that the distribution of mobile protons in the green bean image (Fig. 4A) is denser and more homogeneous compared to the distribution in the roasted bean (Fig. 4B). The moisture content of the green bean was measured using a vacuum oven to be 8.0% moisture, compared to the lower 1.9% moisture in its roasted counterpart. With this knowledge, we hypothesize that the protons in Fig. 4A represent water and lipids, while the protons in Fig. 4B represent mostly mobile lipid protons. Due to the short $T_2$ relaxation time of the coffee bean components and the acquisition parameters used to facilitate comparison of the beans pre- and post-roasting, the lipids and water protons in the green bean were not distinguishable. As expected, image in Fig. 4A appears brighter than Fig. 4B because the roasting process leads to 6.1% loss of moisture.

It is known that lipids exist mostly in the endosperm of a green coffee bean and during bean expansion caused by roasting, the oleosomes are disrupted and lipids migrate to the surface of the bean (Schenker & Rothgeb, 2017). This explains the increase in the yield of oil extraction after roasting. The increase in bean volume or expansion is visually apparent while comparing the small bean size in Fig. 4A to the larger size of the roasted bean in Fig. 4B. It can be concluded that the bright edge around the bean in Fig. 4B represents the mobile lipids that migrated from the endosperm to the surface. This representative result is in agreement with the results obtained using cryo-scanning electron microscopy micrographs (Schenker, Handschin, Frey, Perren, & Escher, 2000). Similar to our work, Schmidt et al. present a proton density image of a coffee bean of 35% moisture which, due to its moisture content, is likely a green bean (Schmidt et al., 1996). To our knowledge, we present the first MR image of a roasted coffee bean which may be a promising tool to investigate oil distribution during roasting of coffee.

**4. Conclusions**

Overall, $^1$H and $^{13}$C NMR spectroscopy proved to be successful tool for the combined targeted and untargeted analysis of coffee lipids. While the $^1$H spectrum offers more sensitive information about the lipid profile of a coffee bean, $^{13}$C was also an efficient tool to quantify individual components. The PULCON method for quantitation was in good agreement with the traditional IS quantification method and can be applied in future studies to prevent sample contamination and other issues related to the addition of an IS in the NMR sample.

Both targeted and untargeted analysis revealed that roasting has a significant impact on the lipid composition of coffee beans. Targeted analysis was able to detect differences in the amount of aldehydes,
terpenes, FFA, LO, and LN following roasting. Additionally, geobraiphical region had a significant impact on K, C, α,β-unsaturated aldehyde, OL, and LO measurements. Both 1H and 13C NMR can be used for untargeted analysis, however, it was found that the higher resolution of 13C NMR does not improve the ability to differentiate between green and roasted coffee lipid samples due to a loss in sensitivity in 13C analysis. Finally, roasting causes changes in the morphology of coffee beans, and MRI is an effective tool to observe the migration of lipids from the endosperm to the surface of the coffee bean.

Acknowledgments

This research was supported by the Department of Food Science and Technology at OSU, the Food for Health Discovery Theme and by the USDA NIFA, Hatch project 1012460. This study made use of the Campus Chemical Instrument Center NMR facility at OSU. Many thanks to Dr. Jiadi Xu at the MRI facility at the Kennedy Krieger Institute at Johns Hopkins University and to Dr. Sravanti Paluri at OSU for providing constructive comments for MRI analysis. Authors are grateful to Professor Devin Peterson for kindly providing coffee samples.

Declaration of Competing Interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2019.125039.

References


