1D and 2D NMR spectra of coffee from 27 countries

Javier Osorio

Chemistry Department, Universidad del Valle, 76008 Cali, Colombia <https://orcid.org/0000-0002-8278-3103>

Victoria A. Arana

Chemistry Program, Basic Science Faculty, Universidad del Atlantico, 081007, Puerto Colombia, Colombia

Jessica M. Medina

Chemistry Program, Basic Science Faculty, Universidad del Atlantico, 081007, Puerto Colombia, Colombia

Rodrigo Alarcon

Almacafé S.A., Bogotá D.C., Colombia

Edgar Moreno

Almacafé S.A., Bogotá D.C., Colombia

Julien Wist (✉ julien.wist@correounivalle.edu.co)

Chemistry Department, Universidad del Valle, 76008 Cali, Colombia <https://orcid.org/0000-0002-3416-2572>

Data Note

Keywords: Magnetic Resonance, Coffee, Origin Determination, Chemometrics

Posted Date: December 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1146879/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Gigabyte on March 11th, 2022. See the published version at <https://doi.org/10.46471/gigabyte.50>.

PAPER

1D and 2D NMR spectra of coffee from 27 countries

Javier Osorio¹, Victoria A. Arana^{2,†}, Jessica M. Medina^{2,†}, Rodrigo Alarcon³, Edgar Moreno³ and Julien Wist^{1,4,*}

¹Chemistry Department, Universidad del Valle, 76008 Cali, Colombia and ²Chemistry Program, Basic Science Faculty, Universidad del Atlantico, 081007, Puerto Colombia, Colombia and ³Almacafé S.A., Bogotá D.C., Colombia and ⁴Australian National Phenome Centre, Computational and Systems Medicine, Health Futures Institute, Murdoch University, Harry Perkins Building, Perth, Australia, WA6150.

*julien.wist@correounivalle.edu.co

†Contributed equally.

Abstract

715 green coffee samples were gathered by Almacafé S.A. (Bogotá, Colombia) between 2012 and 2014 from 27 different countries and analyzed at the Nuclear Magnetic Resonance (NMR) laboratory at Universidad del Valle (Cali, Colombia). Over 1000 methanolic coffee extracts were prepared and a total of 4563 spectra were acquired in a fully automatic manner using a 400 MHz NMR spectrometer (Bruker Biospin, Germany). The dataset spans the variance that could be expected for an industrial application of origin monitoring, including samples from different harvest time, collected over several years and processed by at least two distinct operators. The resulting 1D and 2D spectra can be used to develop and evaluate feature extraction methods, multivariate algorithms and automation monitoring techniques, as a dataset for teaching, or as a reference for new studies on similar samples and approach.

Classifications: Classification 1; Classifications 2; Classifications 3 (Minimum three key words from the GigaScience/GigaDB Classifications List)

Context

In 2005, the Colombian Coffee Federation (FNC, in Spanish) applied for a Protected Geographical Indication (PGI) as an additional move in the branding strategy initiated decades ago with the registration of well-known brands like “Juan Valdez” and “100% Colombian coffee”. The PGI was granted in 2007 and the strategy was met with success as the popularity of Colombian coffees has grown globally. While most coffee were sold as blends, the monitoring of the origin of beans was not a concern. However, during the last two decades, there has been a growing interest for specialty coffees with businesses converting the origin into a major sales argument; thus, creating a need for screening methods to allow for origins of coffee beans to be traced.

In 2010, The Logistic & Quality Control Office of FNC, Almacafé S.A., started screening for analytical platforms for that purpose. At that time, Near InfraRed Spectroscopy (NIRS) was the best known approach, followed by conventional hyphenated techniques, such as Gas or liquid Chromatography coupled to Mass Spectrometry (GC-MS or HPLC-MS) and by Isotope Ratio Mass Spectrometry (IRMS) (See Table 1 in [1]). Little was known about NMR. Two studies reported promising results, one by Consonni et al. [2] using ¹H-NMR and another by Wei [3] and coworkers using ¹³C-NMR, but with small data sets (40 and 60, respectively).

For the following years, Almacafé gathered 715 samples from 27 countries (312), and from most areas producing coffee in Colombia (403) as best represented in Figure 1. At the same time, our laboratory tested and evaluated different protocols for sample preparation and data acquisition. The aim of this collaboration was to evaluate Nuclear Magnetic Resonance (NMR) as

Figure 1. Geographical areas and sample distribution covered or partially covered by this dataset

an analytical platform for large scale monitoring of coffee samples' origin and the major outcomes can be found here.[4]. Later, other techniques were benchmarked in an attempt to find cheaper alternatives to NMR that would produce comparable results, with mIR-ATR (mid-Infrared Attenuated Total Reflectance spectroscopy) being showing promising, with the ability to be portable in the near future.[1, 5] During the acquisition of most of the data, low-field NMR was deemed suitable for a long term study; however; it is a promising technique[6] that should be considered for future research.

The dataset

The dataset presented here was acquired between 2012 and 2014 at the NMR laboratory of the Universidad del Valle (Cali, Colombia). The samples were collected from 27 countries, different species and varieties (See Figure 1, Table 1 and Table 2) by Almacafé S.A (Bogotá, Colombia), during the same period. The total dataset consists of 4563 original folders (with Bruker¹ original hierarchy), two files (tsv format) with descriptions of samples and spectra, and two datasets (tsv format) of 48 and 90 spectra as examples.

Several batches of samples were prepared to answer different research questions with different experimental conditions (See Table 3). Within a batch, all the samples were prepared and the spectra recorded following the same protocol. Four types of NMR experiments were performed as described in Section *Data acquisition* and shown in Table 4. Since the parameter set enforces all other experimental parameters, it suffices to ensure that spectra are compared of same parameter set.²

The resulting spectra for each samples, including replicates, were stored in a folder (See Figure 2 and below).

Folder names and structure

The folder names abide by the convention:

- i. a 6-characters code, 2 capital letters followed by 4 digits that represents the unique identification of each sample. The samples received (715) were named as ACxxxx for Almacafé, while the samples prepared in Cali (mixtures as described in Section *Adulteration*) were named as UVxxxx after Universidad del Valle.
- ii. a section that contains different information on the sample depending on the batch.
- iii. a 7-digits unique identifier that was used by our LIMS[7] (Laboratory Information Management System) to distinguish the several spectra belonging to the same extract, either different types of experiments or repetitions.
- iv. a 10-character key that was used to import data into the database.

The four sections are separated by "_". The second section may contain more than one field, also separated by "_". In order to unequivocally identify the data only the first and the penultimate sections are relevant.

Bruker files are read by several packages[8, 9, 10] and software. Reading and displaying processed Bruker spectra is readily achieved by reading 2 files. The typical folder structure is shown in Figure 2.

Each folder contains a sub-folder for each different experiment performed³.

Each sub-folder contains files and one folder. *acqus* contains information about the experimental conditions for acquisition and *fid* (or *ser* for 2D) contains the actual Free Induction Decay (FID), i.e., the raw data. The data after processing, Fourier transform,

¹ Bruker Biospin is a major manufacturer of NMR spectrometers

² The abbreviation *meoh* in filenames always refers to coffee samples extracted in methanol.

³ By Bruker convention, these folder and sub-folders are referred to as "expname" and "expno"

Table 1. Distribution of the 715 samples by countries and species

Country	Species	Counts
Brazil	Unknown	7
	Arabica	24
	Robusta	4
Cameroon	Robusta	1
China	Arabica	5
	Liberica	1
	Robusta	1
Colombia	Unknown	7
	Arabica	388
	Decaf. Arabica	2
	Catimor	4
	Hibrido de Timor	2
Côte d'Ivoire	Robusta	1
Costa Rica	Arabica	8
Dominican Republic	Arabica	7
	Robusta	6
Ecuador	Unknown	3
	Arabica	4
	Robusta	1
El Salvador	Arabica	4
Ethiopia	Arabica	3
Guatemala	Unknown	3
	Arabica	20
	Robusta	2
Hawaii (USA)	Arabica	2
Honduras	Unknown	7
	Arabica	10
India	Robusta	6
Indonesia	Unknown	1
	Arabica	2
	Robusta	4
Jamaica	Arabica	2
Kenya	Arabica	1
Mexico	Arabica	6
	Robusta	2
Nicaragua	Unknown	1
	Arabica	3
Panama	Arabica	11
Peru	Unknown	3
	Arabica	17
Unknown	Unknown	64
	Arabica	8
	Mixture	18
	Robusta	5
Tanzania	Hibrido de Timor	1
	Unknown	2
	Arabica	1
Togo	Robusta	1
	Arabica	1
Uganda	Robusta	4
	Unknown	3
Venezuela, RB	Arabica	1
	Robusta	17
Vietnam	Decaf. Robusta	1
	Arabica	1
Zambia	Arabica	1

phase-correction and baseline correction, are stored in the `pdata/1/` sub-folder.

Extracting spectra

Five parameters in file `pdata/1/procs/` are needed to read the processed spectrum stored in file `pdata/1/1r` for 1D data (`pdata/1/2rr` for 2D) and display it. Figure 3 shows two typical spectra, from Arabica coffee in black and Robusta coffee in green.

- i. `BYTORDP`: endianness, 0 for *little* and 1 for *big*
- ii. `FTSIZE`: the length of the data buffer

Table 2. Distribution of the 715 samples by species and varieties

Batch	Variety	Counts(1D NOESY)
Unknown	Unknown	99
	EDERWISE	1
	SNOW BERRY	1
Arabica	Unknown	393
	BOURBON	6
	CASTILLO	38
	CATUAI	2
	CATURRA	25
	CATURRA-COLOMBIA	1
	CATURRA-COLOMBIA-CASTILLO	1
	CATURRA(30)-TIPICA(70)	1
	CATURRA(40)-TIPICA(20)-OTRA(40)	1
	CATURRA(70)-CASTILLO(30)	1
	CATURRA(70)-TIPICA(15)-CASTILLO(15)	1
	CATURRA(80)-CASTILLO(20)	5
	CATURRA(90%)-CASTILLO(10%)	1
	COLOMBIA	2
	COLOMBIA (30)- CATURRA (70)	1
	COLOMBIA(3)-CATURRA(90)-OTRA(7)	1
	COLOMBIA(30)-CATURRA(70)	1
	COLOMBIA(30)-TIPICA(50)-BOURBON(20)	1
	COLOMBIA(50)-CASTILLO(50)	1
	COLOMBIA(50)-CATURRA(50)	1
	COLOMBIA(59 %)-CATURRA(41%)	1
	COLOMBIA(65)-CASTILLO(35)	1
	COLOMBIA(70)-CATURRA(30)	1
	COSTA RICA 95	11
	CRUCE	10
	ETIOPE	1
	GEISHA	10
	LAURINA	3
	MANDHELING	1
	MARAGOGIPE	2
	MOGIANA	1
	Ebano Verde	3
Decaf. Arabica	Unknown	2
Liberica	Unknown	1
Mixture	Unknown	18
Robusta	Unknown	51
	CONILON	5
Decaf. Robusta	Unknown	1
Catimor	Unknown	4
Hibrido de Timor	Unknown	3

iii. `NC_proc`: the scaling factor. A value of 3 indicated that the data have been halved 3 times, while negative value indicates that data have been multiplied by two n times.

iv. `SF`: the carrier frequency [MHz]

v. `SW_p`: the spectral width (or the observation window) [Hz]

vi. `OFFSET`: the offset of the origin [ppm]

The x axis in ppm ranges between `OFFSET` and `OFFSET - SWp/SF` with increments of $(SW_p/SF)/(FTSIZE - 1)$.

In addition, useful information about the NMR experimental conditions can be extracted from the `acqus` file contained in each folder.

- `##$EXP= <X>` is the name of the parameter set that ICON NMR, the Bruker Automation Software, called (See Table 4).
- `##$PULPROG= <X>` contains the name of the pulse sequence that was applied to record the spectra (See Table 4).
- `##$RG= X` describes the value of the receiver gain.
- `##$SOLVENT= <X>` is the name solvent parameter set used to lock the frequency and thus gives information about the solvent used for extraction.

Samples description file (metadata)

The file `curatedCodes.tsv` contains all the available information (715 rows and 53 columns) about the samples that were received from Almacafé. Only the `comments` column has been removed to meet privacy concerns. The most important fields, `country`, `department`, `city`, `species`, `variety` and `dateReception` were homogenised and curated, the others were left with original values.

Figure 2. Structure of each folder. Filled circle represents folder, while open circles are for files. Folders with gray circles have similar structure (not represented here for clarity)

Figure 3. Two typical spectra of Arabica (black) and Robusta (green) coffee extracts in non-deuterated methanol. Signals assigned to Kahweol can be used for discrimination. In addition, caffeine content is expected to be higher in Robusta species, as observed in the insert.

Table 3. Number of spectra recorded in each batch

Batch	Counts(all)	Counts(1D NOESY)	Counts(J-RES)
Cafestol	3	1	0
CURVE-roasted	111	30	21
Mix_arabica_meoh_Roasted	64	16	16
Mix_meoh_Roasted	584	146	146
Origin_meoh_green	435	110	105
Origin_meoh_Roasted	1841	461	455
ORIGIN-MeOH	530	134	128
Quantification-green-MeOH	40	10	10
Quantification-roasted-MeOH	64	16	16
Quantification-roasted-MeOH-2	279	70	69
Referencia-Junio-2-2012	3	1	0
Reproducibility_2013	24	6	6
Reproducibility-green-120713	19	6	1
Reproducibility-green-120801	6	2	1
Reproducibility-roasted-120713	37	12	0
Reproducibility-roasted-120801	15	5	0
Reproducibility-roasted-120808	9	3	0
Reproducibility-roasted-MeOH	18	6	0
Roasted-MeOH	64	16	16
roastingEffect	417	105	104

Experiment list

To facilitate the re-use of the data, a file `curatedExtractedNames.tsv` (4563 rows and 23 columns) is provided that contains the path to each experiment (expname/expno) aligned with basic information about the sample, such as: *ID*, *entryID*, *expNo*, *parameterSet*, *pulseProgram*, *solvent*, *baseFrequency*, *frequency*, *temperature*, *spectralWidth*, *receiverGain*, *date*, *path*, *batchID*, *dateReception*, *name*, *country*, *department*, *species*, *variety*, *composition*, *roastingTemperature*, *roastingColor*.

Test dataset

Small datasets, `colombiaOthers48.tsv` and `robustaArabica89.tsv`, are provided as examples or to quickly evaluate the usefulness of the data. The first 3 columns contains the *path*, the *ID* and the *group*. The remaining columns contain the spectra trimmed from 0.01 to 10 ppm and after removal of the methanol region between 3.15 and 3.6 ppm. The column headers are the ppm values.

Methods

Several protocols and data analysis pipelines were evaluated and several research questions were raised as our curiosity was piqued by preliminary results. Some of them were explored as reflected by the inhomogeneous granularity of the metadata (`curatedCodes.tsv`) and the large number of reported batches (See Table 3).

Sample recollection

Coffee samples (*Coffea arabica*, NCBI:txid13443 and *Coffea canephora* var. *Robusta*, NCBI:txid308126) were recollected by the Laboratory for Quality Control at Almacafé according to availability. This laboratory is in charge of monitoring the quality of the coffee beans during all the processes that take place prior to exportation. As sole contractor of the FNC, Almacafé must monitor all the coffee harvested and exported by the FNC and therefore has access to coffee samples from all regions of Colombia. Furthermore, as major producer and worldwide exporter, the FNC and Almacafé have access to international partners and it was thus able to gather samples from other producing countries.

Sample preparation for NMR

Obtaining the fastest, most robust, and cheapest protocol to prepare the samples was challenging. This included performing solid-liquid extraction at ambient temperature (to avoid controlling an additional parameter) and in non-deuterated solvent (to lower the cost per experiment). Best results were obtained using centrifugation to remove solid particles[4] after extraction. Filtration, considered initially, was discarded due to poor reproducibility (See Table 3, Reproducibility-*).

Different solvents were compared for extraction, water, chloroform and methanol. Namely best information (number of signals) was obtained from methanol extracts, and thus methanol was chosen as sole solvent for the rest of the project. However, chloroform produced interesting spectra, extracting apolar compounds, such as fatty acids, that were later found important for origin determination.

Both green and roasted beans were considered. Ideally, green coffee samples (Table 3: *Origin_meoh_green*) would have been preferred, since it avoids the process of roasting. However, this latter process was already well implemented and extensively used by Almacafé, while the grinding of green beans required ball grinders not widely available at the time. This obstacle could

Table 4. Number of spectra recorded grouped by pulse sequence

Name of experiment	Name of parameter set	Pulse sequence file name	Dimension	Counts
J-Resolve[11]	COFFEE_meoh_jres4.lims	jresgpps1f.2	2D	1094
1D NOESY[12]	COFFEE_meoh_noesyp64.lims	noesygpps1d.comp	1D	1156
¹ H-NMR	COFFEE_meoh_pilot1	zg30	1D	1157
¹ H-NMR with pre-saturation[13]	COFFEE_meoh_pilot2	zgpsdo	1D	1156

have been overcome, but green coffee extracts were showing rapid oxidation while waiting for acquisition. And more importantly, samples to be analyzed from the market will most likely already be roasted. Therefore, the samples were roasted according to the standard procedure applied for cupping and ground to a very fine powder (Table 3: *Origin_meoh_Roasted*).

In addition to regular solution-state ¹H-NMR were evaluated, other techniques, such as Solid-state MAS (Magic Angle Spinning), HR-MAS (High-Resolution Magic Angle Spinning). The latter two techniques present the advantage that no solid-liquid extraction is needed, reducing contamination, saving time and money. Since these spectra were recorded on different machines kindly made available by Bruker Biospin, which are not listed here.

Data acquisition

NMR allows us to observe nuclear spins, in particular those of protons. Therefore chemists are used to dissolve the substance of interest in deuterated solvents, i.e., solvents where all protons have been replaced by deuterons that are invisible to NMR. The result is that only the molecules of interest will contribute to the signal and thus to the spectra. Since the solvent is always much more concentrated than the substrate, its signal will obfuscate the signal of interest, in the same way as the sun makes stars invisible during the day. Deuterated solvents are much more expensive and thus less suitable for large scale screening, in particular in developing economies. Therefore, the extracts were prepared using regular methanol (chromatographic grade), while the spectra were acquired using a solvent signal suppression scheme[13, 14, 15]. To achieve a robust and completely automatic (without human supervision) suppression of the solvent signal, a row of 3 experiments was performed for each sample, (*COFFEE_meoh_pilot1*) to estimate the resonance frequency of the solvent, (*COFFEE_meoh_pilot2*) to determine this frequency with more accuracy by applying a light suppression and (*COFFEE_meoh_noesyp64.lims*) to acquire the spectra by suppressing the solvent at that frequency[12]. For some samples, an additional 2D experiment was performed *COFFEE_meoh_jres4.lims*⁴. J-Resolved experiments provide a wealth of information about scalar couplings constants[11] and have been less exploited due to their size and because peak-picking and peak-fitting are more complex to perform in 2D.

An important technical note: during the data acquisition, the spectrometer experienced a technical issue that modified the carrier frequency for the subsequent experiments. As such, the data was re-acquired for some of the samples for the sake of publication, and it also provided a unique opportunity to assess how carrier frequency can effect the field stability of the final models. The information about the carrier frequency can be found for each individual spectra, in its parameter file *acqus* under the name *SFO1*. Before the incident the frequency was of 400.13188235 MHz, and it stabilized to 400.02188249412 MHz, after the ice was removed and the pressure in the Dewar returned to normal. For simplicity, this information has been included in the file *curatedExtractedNames.tsv* in the column *baseFrequency*.

Complete details about the sample preparation and data acquisition can be found elsewhere[4] and within the original files, in particular in the *acqus* files available for each experiment.

Origin determination

The determination of coffee origin is the major focus of Almacafé and, represents the core of this study, thus; the largest share of the data, with more than 2000 acquired spectra (Table 3: *Origin_meoh_green*, *Origin_meoh_Roasted* and *ORIGIN_MeOH*). A large effort was dedicated to covering a vast geographical area. Special care was taken to include samples from neighboring areas that are expected to share most similarities. When available, information about the post-harvest processing, such as washing and drying was recorded, as it is expected to strongly affect the final product. In addition, the numerous samples available from Colombia allows one to look into origin by department or by region with similar agro-climatological characteristics. For a quick look into this batch, a test dataset *colombiaOthers48.tsv* was prepared by randomly picking 24 samples from Colombia and 24 from alternate countries.

Adulteration

A second interest of the Laboratory for Quality Control of Almacafé is the possible adulteration of Colombian coffee with lower quality Robusta (*Coffea canephora* var. Robusta) beans imported to Colombia for internal consumption. Even before origin, the coffee species have been extensively used for marketing. A large body of literature has been dedicated to detect such kind of adulteration[16] and it was deemed to be a natural starting point for comparison with our results. Therefore, 252 samples (Table 3: *Mix_arabica_meoh_Roasted*, *Mix_meoh_Roasted*, *Quantification-roasted-MeOH* and *Quantification-roasted-MeOH-2*) were prepared by adulterating Arabica coffee with concentrations of Robusta ranging from 5% to 90%. Colombia only produces coffee Arabica. Since the species is expected to have a larger effect size than the origin, a first classification may be performed to exclude all non-Arabica samples (including adulterated ones) prior to modeling the origin[4]. Cafestol is a biomarker for discriminating Robusta

⁴ 2D experiments encode information about chemical shifts and scalar couplings in 2 dimensions, to highlight correlations and reduce peak overlap

from Arabica. [17, 18] A reference spectrum is shared that was prepared in similar conditions and can be used to help identify signals. For a quick look into this batch, a test dataset `robustaArabica89.tsv` was prepared by randomly picking 45 samples from Colombia and 45 from alternate countries. It should be noted, one sample with poor shimming was removed.

Effect of roasting

During this collaboration, the authentication of several commercial samples was conducted by Almacafé from retailers around the world. Some of these samples are listed here as *Unknown*. The samples were first flagged as suspicious by cupping and then sent for analysis. Although the classification performed well in most cases, some samples were found to be difficult to classify⁵. These latter samples raised the question of how the roasting process could confuse the models. Indeed, the samples were roasted by the retailers, and not according to the standard process used to build the classifier. A test was thus conducted (See Table 3: `RoastingEffect` and `CURVE-roasted`) to estimate the impact of roasting temperature and final color[19] on the resulting profiles (spectra). It was proven possible to contain this effect by excluding the features of the spectra most affected by the degree of roasting (See chapter 3.5 of [20]), although, beyond some color (very dark) samples could not be modeled accurately.

Replicates

Only considering experiments from `Origin_meoh_Roasted` batch and acquired with `noesygpps1d.comp`, more than 83 (out of 353) samples were acquired with replicates, either to monitor the long term stability of the process and to estimate its associated experimental noise, or because the automation failed leading to poor quality spectra. All these spectra are shared here and can be used to develop and test algorithms that are capable of detecting failed acquisition and to detect batch or run effects, using the acquisition time available in the `acqus` files.

Data Validation and Quality Control

The data presented here were recorded during a collaboration project between the Laboratory of Quality Control of Almacafé S.A. (Bogotá, Colombia) and the NMR Laboratory at Universidad del Valle (Cali, Colombia). Some of the outcomes of this project can be found here [4, 1, 5], along with more details about the protocols used for sample preparation and about the experimental setup to acquire the data.

Reuse Potential

To the extend of our knowledge this dataset represents the largest available database of NMR spectra of coffee samples, and as such could be used as a reference baseline for future studies and to benchmark data reduction methods or classification algorithms. The size and complexity (a large amount of sparse metadata covering numerous degrees of freedom) of the dataset makes it a useful resource for teaching. This motivated us to release the complete dataset, including spectra acquired during the validation of the methodology.

Availability of supporting data and materials

The dataset supporting the results of this article is available here: <https://doi.org/10.5281/zenodo.4126117>. For any further details, or complementary information, please contact the corresponding author.

Declarations

List of abbreviations

NMR; LIMS; NIRS; GC-MS; HPLC-MS; IRMS; NOESY; FID, FNC; PGI

Consent for publication

Not applicable

Competing Interests

The authors declare no competing interests.

⁵ Among them were found several decaffeinated samples that confused the cupping

Funding

This research was funded by Almacafé S.A., Bogotá, Colombia.

Author's Contributions

Conceptualization, JW, RA, EM; Data acquisition, VA and JM, Data Curation, JW and JO; Project funding and supervision, JW; Data analysis, VA, JM, JO and JW; Drafting the manuscript, JW; Revision and edition, JO and VA;

Acknowledgements

Not applicable

References

1. Arana VA, Medina J, Esseiva P, Pazos D, Wist J. Classification of Coffee Beans by GC-C-IRMS, GC-MS, and ¹H-NMR. *J Anal Methods Chem* 2016;2016.
2. Consonni R, Cagliani LR, Cogliati C. NMR based geographical characterization of roasted coffee. *Talanta* 2011 Nov;.
3. Wei F, Furihata K, Hu F, Miyakawa T, Tanokura M. Complex mixture analysis of organic compounds in green coffee bean extract by two-dimensional NMR spectroscopy. *Magn Reson Chem* 2010;48(11):857–865.
4. Arana Va, Medina J, Alarcon R, Moreno E, Heintz L, Schäfer H, et al. Coffee's country of origin determined by NMR: The Colombian case. *Food Chem* 2015;175:500–506.
5. Medina J, Rodríguez DC, Arana VA, Bernal A, Esseiva P, Wist J. Comparison of Attenuated Total Reflectance Mid-Infrared , Near Infrared , and ¹H-Nuclear Magnetic Resonance Spectroscopies for the Determination of Coffee ' s Geographical Origin. *Int J Anal Chem* 2017;2017.
6. Defernez M, Wren E, Watson AD, Gunning Y, Colquhoun IJ, Le G, et al. Low-field ¹H NMR spectroscopy for distinguishing between arabica and robusta ground roast coffees. *Food Chem* 2017;216:106–113.
7. Patiny L, Zasso M, Kostro D, Bernal A, Castillo AM, Bolaños A, et al. The C6H6 NMR repository: an integral solution to control the flow of your data from the magnet to the public. *Magn Reson Chem* 2017;.
8. Kimhofer T. MetaboMate: All you need for NMR-based metabolic profiling with R!; 2020, <https://github.com/tkimhofer/MetaboMate>, r package version 0.0.0.9002.
9. Helmus JJ, Jaroniec CP. NmrGlue: an open source Python package for the analysis of multidimensional NMR data. *J Biomol NMR* 2013 Apr;55(4):355–367.
10. Organisation TC, Cheminfo.org; 2020. Accessed: 2020–10–23. <http://www.cheminfo.org/flavor/metabolomics/index.html>.
11. Aue WP, Karhan J, Ernst RR. Homonuclear broad band decoupling and two-dimensional J-resolved NMR spectroscopy. *J Chem Phys* 1976 May;64(10):4226–4227.
12. McKay RT. How the 1D-NOESY Suppresses Solvent Signal in Metabonomics NMR Spectroscopy: An Examination of the Pulse Sequence Components and Evolution. *Concepts Magn Reson* 2011;38:197–220.
13. Hoult DI. Solvent peak saturation with single phase and quadrature fourier transformation. *J Magn Reson* 1976 Feb;21(2):337–347.
14. Price WS. Water signal suppression in NMR spectroscopy. *Annu Rep NMR Spectrosc* 1999;38.
15. McKay RT. Chapter 2 Recent Advances in Solvent Suppression for Solution NMR: A Practical Reference. In: *Annual Reports on NMR Spectroscopy*, vol. 66 Academic Press; 2009.p. 33–76.
16. Toci AT, Farah A, Pezza HR, Pezza L. Coffee Adulteration: More than Two Decades of Research. *Crit Rev Anal Chem* 2016;46(2):83–92.
17. Cavin C, Holzhaeuser D, Scharf G, Constable A, Huber WW, Schilter B. Cafestol and kahweol, two coffee specific diterpenes with anticarcinogenic activity. *Food Chem Toxicol* 2002 Aug;40(8):1155–1163.
18. D'Amelio N, De Angelis E, Navarini L, Schievano E, Mammi S. Green coffee oil analysis by high-resolution nuclear magnetic resonance spectroscopy. *Talanta* 2013 Jun;110:118–127.
19. Leme DS, da Silva SA, Barbosa BHG, Borém FM, Pereira RGFA. Recognition of coffee roasting degree using a computer vision system. *Comput Electron Agric* 2019 Jan;156:312–317.
20. Arana VA. Autenticación del origen botánico y geográfico del café colombiano por metabolómica basada en RMN. PhD thesis, Universidad del Valle; 2015.