Identification of biochemical features of defective Coffea arabica L. beans

María I. Casas, Michael J. Vaughan, Pierluigi Bonello, Brian McSpadden Gardener, Erich Grotewold, Ana P. Alonso

PII: S0963-9969(17)30077-7
DOI: doi: 10.1016/j.foodres.2017.02.015
Reference: FRIN 6598

To appear in: *Food Research International*

Received date: 19 August 2016
Revised date: 22 February 2017
Accepted date: 26 February 2017

Please cite this article as: María I. Casas, Michael J. Vaughan, Pierluigi Bonello, Brian McSpadden Gardener, Erich Grotewold, Ana P. Alonso , Identification of biochemical features of defective Coffea arabica L. beans. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Frin(2017), doi: 10.1016/j.foodres.2017.02.015

TITLE: Identification of biochemical features of defective *Coffea arabica* L. beans

AUTHORS: María I. Casas[a,b,c,1]; Michael J. Vaughan[b,d,1]; Pierluigi Bonello[b,e]; Brian McSpadden Gardener[b,e] ; Erich Grotewold[b,c] and Ana P. Alonso[b,c,*]

AFFILIATIONS:

[a]Molecular Cellular and Developmental Biology Graduate Program, The Ohio State University, 1060 Carmack Road, Columbus, OH, 43210, USA

[b]Center for Applied Plant Sciences (CAPS), The Ohio State University, 1060 Carmack Road, Columbus, OH, 43210, USA

[c]Department of Molecular Genetics, The Ohio State University, 1060 Carmack Road, Columbus, OH 43210, USA.

[d]Department of Plant Pathology, The Ohio State University-OARDC, 1680 Madison Ave.,Wooster, OH, 44691.

[e]Department of Plant Pathology, The Ohio State University, 2021 Coffey Road, Columbus, OH 43210, USA.

[1] The two authors contributed equally to this work and should be considered as co-first authors.

[*]Corresponding author: Dr. Ana P. Alonso, The Ohio State University, Department of Molecular Genetics, 054 Rightmire Hall, 1060 Carmack Road, Columbus, OH, USA. Email: alonso.19@osu.edu; Ph: +001-614-688-7404; Fax: +001-614-247-8937.

ABSTRACT

Coffee organoleptic properties are based in part on the quality and chemical composition of coffee beans. The presence of defective beans during processing and roasting contribute to off flavors and reduce overall cup quality. A multipronged approach was undertaken to identify specific biochemical markers for defective beans. To this end, beans were split into defective and non-defective fractions and biochemically profiled in both green and roasted states. A set of 17 compounds in green beans, including organic acids, amino acids and reducing sugars; and 35 compounds in roasted beans, dominated by volatile compounds, organic acids, sugars and sugar alcohols, were sufficient to separate the defective and non-defective fractions. Unsorted coffee was examined for the presence of the biochemical markers to test their utility in detecting defective beans. Although the green coffee marker compounds were found in all fractions, three of the roasted coffee marker compounds (1-methylpyrrole, 5-methyl- 2-furfurylfuran, and 2-methylfuran) were uniquely present in defective fractions.

KEYWORDS: coffee; flavor; volatiles; amino acids; sugars; organic acids

CHEMICAL COMPOUNDS STUDIED IN THIS ARTICLE: 1H-pyrrole, 1-methyl- (PubChem CID: 7304); 2-(2-furanylmethyl)-5-methyl- (PubChem CID: 595524); 3-methylbutanal (PubChem CID: 11552); acrylic acid (PubChem CID: 6581); citric acid (PubChem CID: 311); fructose (PubChem CID: 5984); malic acid (PubChem CID: 525); pentitol (PubChem CID: 827); pyridine (PubChem CID: 1049); sucrose (PubChem CID: 5988).

1. Introduction

There are two main coffee species with commercial value, *Coffea canephora* Pierre ex A. Froehener (also known as Robusta coffee) and *Coffea arabica* L. (known as Arabica coffee), which are an important commodity crop produced in tropical countries. Once beans are roasted, they are used to prepare the coffee beverage. Arabica coffee is characterized by a superior flavor and aroma qualities (Flament, 2002). The market value and use of a given coffee depends on its organoleptic properties. Therefore, determining the elements influencing coffee quality remains an important field of study. Many factors, including chemical composition, the presence of defects, and the roasting process can impact the final flavor and aroma of coffee (Franca, Mendoca, & Oliveira, 2005; Mazzafera, 1999).

Before roasting, coffee is traded as dried seed, known as green coffee. Metabolites accumulating in green coffee beans, act as precursors for the compounds responsible for flavor and aroma after roasting (George, Ramalakshmi, & Rao, 2008). The chemical composition of green coffee beans is not only dependent upon cultivar, but is also affected by the terroir, harvesting methods (e.g. hand picked or mechanical), seed processing (e.g. wet, dry, or semi dry) and storage (Franca, Oliveira, Mendoca, & Silva, 2005; Illy & Viani, 2005). Coffee producers often try to exploit this knowledge to achieve higher quality coffee. For instance, coffee processing is usually started the same day that fruits are harvested to prevent unwanted fermentation that could alter bean chemistry. Additionally, coffee seeds are routinely graded and sorted based on bean size, color and shape to reduce the presence of defects and ensure uniform roasting (Batista & Chalfoun, 2014; Franca, Oliveira et al., 2005; Illy & Viani, 2005). The main classes of

metabolites present in green coffee beans have been identified as amino acids, carbohydrates, organic acids, fatty acids, chlorogenic acids (CGAs), and volatile organic compounds (Farah, 2012). Several metabolites accumulating in green beans contribute to flavor and aroma after roasting. Caffeine, for instance, is one of the highest accumulating purines in green coffee beans, and is partly responsible for the bitter taste of coffee beverage (Farah, 2012). The sugar content in coffee beans, particularly sucrose, provides critical precursors for flavor production during the roasting process (Farah, 2012). Roasting generates numerous soluble and volatile compounds responsible for flavor and aroma. It is a complex chemical process involving Maillard and Strecker reactions during which amino acids, sugars, trigonelline and CGAs are converted to furans, pyrazines, pyridines, and other nitrogen- and sulfur-containing metabolites responsible for many of the organoleptic properties of the coffee beverage (Franca, Mendoca et al., 2005; van Boekel, 2006).

Coffee bean defects include irregular seed shape, insect damage, over-fermentation, and seed discoloration (Illy & Viani, 2005). Defective coffee beans reduce cup quality by altering the organoleptic properties of their roasted counterparts (Flament, 2002). Characterizing the chemical composition of these beans is essential to identify the metabolites responsible for off-flavors. The proportion of defective beans can exceed 20% by weight in the case of some coffee consumed in Brazil and in other producing countries around the world (Franca, Oliveira et al., 2005; Mendoca, Franca, Oliveira, & Nunes, 2008; Ramalakshmi, Kubra, & Rao, 2007). Previous studies reported changes in composition of CGAs and volatiles in defective green coffee beans, which affect the composition of roasted seeds and, as a consequence, coffee cup quality (Mazzafera, 1999;

Oliveira, Franca, Mendoca, & Barros-Junior, 2006; Ramalakshmi et al., 2007). Those studies however, focused either on a very particular defect or a specific class of compounds, not taking into account the contributions of such defects in the whole pool of beans from which they originated. For instance, 2-methylpyrazine and 2-furylmethanol acetate accumulated mainly in green immature defective seeds, whereas butyrolactone was mostly found in green sour seeds. Benzaldehyde and 2,3,5,6-tetramethylpyrazine in green coffee might be associated with black-immature defective beans (Toci & Farah, 2008).

The overall goal of this study is to better understand the impact defective beans have on coffee quality, and more specifically to identify a set of metabolites that are indicative of the presence of defects. To this end, defective and non-defective samples were collected from an average commercial batch of coffee and subsequently analyzed as both green and roasted seeds (Figure 1 A and B). The experimental objectives of this work were to i) compare the quantity of the main family of analytes associated with coffee aroma and flavor in defective and non-defective seeds and ii) identify particular compounds that could be used as specific biochemical markers of defective seeds.

2. Materials and methods

2.1. Plant materials and sample processing

Green *C. arabica* beans from Sumatra (Indonesia) were obtained from Sweet Maria's green coffee supplier (www.sweetmarias.com) and stored in the dark at room temperature. Both batches of coffee were from northern Sumatra, the first batch classified as a Mandheling and the second as a Lingtong. Prior to processing, half of the beans from

each coffee batch were hand sorted into defective and non-defective lots. Defective beans were defined as those expressing visible physical deformities (such as pea berry, withered seeds, and shells), seed damage (due to processing or insect damage), or color anomalies (such as full black, partial black, or bleached seeds) based on the parameters set out for seed defects by the SCAA (Kosalos, Stephen, & Diaz, 2013) (Figure 1A). Sorting was conducted in standard laboratory conditions under ambient lighting. Individual seed defects were pooled to form a single defective fraction for analysis. Portions of these coffee seed lots, unsorted, defective, and non-defective, were separated for biochemical profiling and roasting. Approximately 100 g of each green bean fraction were roasted to color on a Nesco model CR-1000 Professional Coffee Roaster (The Metal Ware Corporation, Two Rivers, WI, USA). Coffee seeds were roasted to a target lightness value of 21.5 ± 0.5 L*, corresponding to a light medium roast matching the Specialty Coffee Association of America (SCAA) tile No. 65 in the SCAA Roast Color Classification System (SCAA, Long Beach, CA 90802). To achieve the target roast degree, the roast cycle was set for 15 minutes, reaching a maximum temperature of 206 ± 4 °C. After roasting, coffee samples were allowed to cool for five minutes at room temperature, and then 12 g of each sample were ground in a coffee grinder (Model 80365, Hamilton Beach, NC, USA) using the percolator size setting for four cups. Lightness values of ground coffee were measured using a MiniScan EZ (Reston, Virginia, 20190) following the manufacturer instructions. Roasted samples not matching the target roast degree were discarded and another 100 g sample of green beans was roasted until the target color was achieved. Coffee samples roasted to the correct color were stored in airtight containers at -20°C until further analysis.

2.2. Chemicals

Internal standards for each analysis were obtained from Sigma Aldrich (www.sigmaaldrich.com) and included glyceryl triheptadecanoate (C:17-TAG) for fatty acid quantification; 2-phenyl butyric acid for organic acid analysis; [U-$^{13}C_2$]glycine and [U-$^{13}C_6$]glucose for amino acid and sugars/sugar alcohols quantifications respectively; and 4-heptanone for volatile analysis.

2.3. Determination of physical and chemical properties

Percent water content of green defective and non-defective beans was determined gravimetrically after freeze-drying to maintain bean structure. Briefly, green beans from each category were placed in a 45 mL capacity tube with a perforated lid and freeze-dried until constant weight. For a 100-seed weight determination, one hundred defective and non-defective green coffee beans was counted and weighed on an analytical scale (Sartorius, Bohemia, NY, USA). Initial pH and titratable acidity (TA) were determined on ground samples using 7.5 g of ground beans extracted with 125 mL of boiling water. The ground coffee was removed by straining after 3 min of infusion, with the resulting brews being set aside to cool to room temperature. Then, 101 g of each coffee brew was placed on an automatic titration meter (Mettler-Toledo, Columbus, OH, 43240) to measure its initial pH, equilibration point, and the milliliters of titrant (0.1 M NaOH) needed to reach each of those points. An aliquot of the same brew, ca. 500 µL, was used to determine the percentage of brew solids before titration using a refractometer. Titratable acidity was then calculated as milliliters of titrant needed to reach equilibration

point divided by the % brew solids. Three biological replicates were used in all experiments.

2.4. Analysis of volatile organic compounds

Volatile organic compounds (VOCs) were analyzed by sampling the headspace of 3 g of intact green or roasted beans placed in 20 mL glass vials (Chromacol, Thermo Fisher, Waltham, MA, USA). The difference in volume occupied by roasted and green coffee beans was corrected by adding 5 mm glass beads (Thermo Fisher), to visually reach the same solid levels between green and roasted samples. Then, 1 μL of a 1:100 dilution in methanol of 4-heptanone (98% purity, internal standard) was added to each vial, which was immediately sealed. Samples were incubated at 100 °C for 40 min under a 10 s on/off agitation cycle using a TriPlus TSH autosampler (Thermo Fisher). One mL of headspace volatiles was injected and analyzed using a Thermo Trace 1310 gas chromatograph coupled to a single quadrupole ISQ mass spectrometer (Thermo Fisher). VOCs were separated with an Omegawax Capillary GC Column (L × I.D. 30 m × 0.25 mm, 0.25 μm film thickness, Supelco). Initial temperature was set at 40 °C, and increased to 230 °C at 4 °C/min, and finally increased to 280 °C at 50 °C/min and held for 2 min. Helium was used as carrier gas at 1.2 mL/min and the injection mode was set to split-less for 0.25 min. The injection temperature was set at 260 °C. Ion source and transfer line temperatures were set at 275 °C and 280 °C, respectively, with electron ionization (EI) mode at 70 eV. Detection was in full scan mode from m/z 40 to 400 with a dwell time of 0.075 s. VOCs were identified using the NIST library (NIST v.11; www.nist.gov) and

amounts were expressed as area ratio to the internal standard (4-heptanone) per gram of fresh weight.

2.5. Extraction of water-soluble metabolites: amino acids, organic acids, sugars and sugar alcohols

Fifteen mg of dried sample were extracted with boiling water as previously described (Cocuron, Anderson, Boyd, & Alonso, 2014), with the following modifications: 10 μL of a 5 mM [U-$^{13}C_2$]glycine - [U-$^{13}C_6$]glucose mix and 30 μL of a 1 mg/mL 2-phenylbutyric acid solution were added as internal standards for amino acids, sugars/sugar alcohols, and organic acids, respectively. After overnight lyophilization, all the extracts were re-suspended in 1 mL H_2O of which 500 μL were loaded onto a 3kDa-Amicon filter and centrifuged for 45 min at 17,000 g and 4 °C. The filtered extract was used to quantify the amino acids, sugars/sugar alcohols, and organic acids (Cocuron et al., 2014; Koubaa, Cocuron, Thomasset, & Alonso, 2013).

2.6. Amino acids quantification

Filtered extracts were diluted 50 times in 0.01 N HCl, then 5 μL or 10 μL of green and roasted extracts were analyzed by LC-MS/MS, respectively, for the quantification of the following amino acids: alanine, arginine, asparagine, aspartic acid (aspartate), cysteine, glutamic acid (glutamate), glutamine, glycine, histidine, isoleucine, leucine, lysine, methionine, hydroxyproline, phenylalanine, proline, serine, threonine, tryptophan, tyrosine and valine (Supplementary Table 2). Amino acids separation and detection were performed using a Hypercarb column (100 x 2.1 mm, 5 μm pore) in an

Agilent 1290 UPLC system, followed by tandem mass spectrometry using a QTRAP 5500 hybrid triple quadrupole/ion trap (AB Sciex), as previously described (Cocuron et al., 2014). LC-MS/MS data were acquired and processed using Analyst software version 1.6.1. Each amino acid was quantified in nmol /mg dry weight using [U-$^{13}C_2$]glycine as internal standard and external standard curves as previously described (Cocuron et al., 2014).

2.7. Quantification of sugars and sugar alcohols

Sugars and sugar-alcohols filtered extracts from green and roasted beans were diluted 50 and 20 times using a mix of acetonitrile:H_2O (60:40, v:v) and then 15 µL and 20 µL were analyzed by LC-MS/MS, respectively, for the quantification of the following sugars and sugar alcohols: fructose, glucose, hexitols, inositol, pentitols, sorbitol, sucrose and tetraols (Supplementary Table 2). The separation and quantification of sugars and sugar alcohols were performed on the same instruments as mentioned above. An Asahipak NH2P-50 2D (2.0 x 150 mm, 5 µm pore) column with an Asahipak NH2P-50G 2A guard column (Shodex) were used to separate sugars and sugar alcohols as previously described (Cocuron et al., 2014). LC-MS/MS data were acquired and processed using Analyst version 1.6.1. Each sugar and sugar alcohol was quantified in nmol/mg dry weight using [U-$^{13}C_6$]glucose as internal standard and external standard curves as previously described (Cocuron et al., 2014).

2.8. Organic acid analysis

Five hundred μL of previously filtered sample were transferred to a new glass tube, frozen in liquid nitrogen and lyophilized overnight. Dried samples were re-suspended in 500 μL of acetonitrile:N,O-Bis(trimethylsilyl)trifluoroacetamide + 1% trimethylsilyl chloride (BSTFA + 1% TMCS) (1:1) and incubated for 1 hr at 120 °C. Derivatized samples were centrifuged for 1 min at 560 g and the supernatant was transferred to 2 mL GC-MS vials.

GC-MS was selected for the analysis of organic acids to enable the use of the mass library from the National Institute of Science and Technology (NIST v.11; www.nist.gov) to identify a wide range of compounds that might not be commercially available as standards. One μL of each sample was injected and analyzed using the same instrument as for headspace analysis. The GC-MS separation was performed with a TG-5MS capillary column (5% phenyl / 95% dimethylpolysiloxane, L x I.D. 30 m x 0.25 mm, 0.5 μm film thickness, Thermo). GC conditions were as follows: initial temperature was set at 100 °C and held for 0.5 min, then increased to 260 °C at 4 °C/min, and finally to 325 °C at 50 °C/min and held for 5 min. The helium carrier gas flow and the split were 1.4 mL/min and 20, respectively. The injection temperature was set at 300 °C. MS conditions included ion source and transfer line temperatures set at 300 °C and 325 °C respectively, with electron ionization (EI) mode set at 70 eV. Detection was in full scan mode between m/z 75 and 650 with a dwell time of 0.084 s (Koubaa et al., 2013). Relative amounts of organic acids were expressed as the ratio to internal standard (2-phenylbutyric acid) per mg of dry weight.

2.9. Fatty acid extraction and analysis

Fifteen mg of ground samples were extracted with 950 µL isopropanol:hexanes (1:2, v:v) and 50 µL of a 10 mg/mL glyceryl triheptadecanoate solution as internal standard. Samples were homogenized with a 5 mm diameter tungsten bead and a bead-beater for 5 min using a (Restch MM400) at a 30 Hz frequency. Oil was extracted and then trans-esterified into fatty acid methyl esters (FAMEs) as previously described (Cocuron et al., 2014) with the following modifications: i) 300 µL toluene and 1 mL of 3N methanolic HCl were used for derivatization; ii) 500 µL 5% w/v $NaHSO_4$ were added to quench the reaction; and iii) 1.7 mL hexanes was supplemented at the end of the derivatization. FAMEs were further centrifuged for 2 min at 560 g and transferred to a new glass tube previously rinsed with hexanes. FAMEs were finally diluted 5 times in hexanes and 1 µl was injected on the GC-MS for quantification of the following fatty acids: arachidic, behenic, eicosenoic, linoleic, linolenic, oleic, palmitic and stearic acids (Supplementary Table 2).

GC-MS analysis of FAMEs was performed on the same instrument as mentioned above using an Omegawax Capillary GC Column (L × I.D. 30 m × 0.25 mm, 0.25 µm film thickness, Supelco). Briefly, initial temperature was set to 170 °C and held for 0.5 min, then increased to 245 °C at 100 °C/min and held for 5.25 min. Helium was used as carrier gas with a flow of 1.4 mL/min and the split was set to 10. The injection temperature was set to 225 °C. For MS conditions, the ion source and transfer line temperatures were set to 200 °C and 250 °C respectively, with electron ionization (EI) mode of 70 eV. Detection was in full scan mode between m/z 20 and 400 with a dwell time of 0.075 s. GC-MS data were acquired and processed using Xcalibur software. FAMEs were identified using the NIST 11 library and FAME standards purchased from

Sigma. Fatty acids were quantified according to the internal standard (triheptadecanoin) and expressed in μg/mg dry weight.

2.10. Statistical analyses and identification of markers for defective seeds

Chemical analyses were conducted on unsorted, non-defective, and defective coffee fractions for both green and roasted beans (n=4 for each category for both green and roasted states). The experiments were conducted on two independent batches of Sumatra coffee. Initial comparisons between defective and non-defective beans, in both green and roasted samples, were performed using two-tailed, type 3, Student's test (*t*-test) considering statistically significant *p* values below 0.05 using MetaboAnalyst version 3.0 (www.metaboanalyst.ca) (Xia, Mandal, Silenikov, Broadhurst, & Wishart, 2012; Xia, Psychogios, Young, & Wishart, 2009; Xia, Silenikov, Han, & Wishart, 2015). Subsequently, heat maps were generated to examine differential compound enrichment between defective and non-defective beans. Briefly, metabolite levels were first transformed using log2 function, and then mean-centered and divided by the standard deviation of each variable. Finally, Ward's hierarchical clustering algorithm was used to group metabolites that have a similar pattern of distribution.

Multivariate tests of association were conducted to identify which compounds might act as indicators of defective beans in a batch of coffee. The analysis was conducted in two stages: i) data reduction and identification of candidate marker compounds and ii) validation of marker compounds using the second independent batch of coffee beans. First, a concatenated data set composed of all metabolite classes analyzed for the first coffee batch was subjected to serial-reductive principal component

analysis (PCA) using variable loadings. Compounds with more than 75% missing data were excluded. Analyses for green and roasted coffee were conducted separately. After having identified a core set of possible marker compounds, the presence of the core markers identified in batch 1 was validated in batch 2 as described above. Ordinations were conducted using Minitab software version 16 (Minitab Inc, PA, USA).

3. Results and discussion

3.1. Determination of bean physical and chemical properties

Coffee defects encountered in this study included irregular shape, insect damage, and black discoloration (Figure 1A). Defective beans constituted approximately 7% (w/w) of the assayed batch. The proportion of each defect was not measured in this study since our main goal was to identify which markers were quantifiable within a pool of beans. To determine how the presence of these defects affected the physico-chemical properties of the beans, water content, 100-seed weight, pH and titratable acidity were measured for defective and non-defective samples (Supplementary Table 1). Water content and 100-seed weight were found to be significantly lower ($p < 0.01$) in defective compared to non-defective seeds. This could be due to the loss of physical integrity of the broken or malformed beans present in the defective fraction. These results are in accordance with previous work on immature and green-immature defects (Mazzafera, 1999). The remainder of the physical characteristics for green coffee was not significantly different between the two fractions, indicating that pH and titratable acidity were not impacted in the defective pool in this study. No significant differences were

observed for roasted samples (Supplementary Table 1), suggesting that any differences in mass or water content were eliminated during the roasting process.

3.2. Metabolite profiling

In this study, we have identified and quantified a total of 119 individual components in green and roasted coffee beans following the workflow schema presented in Figure 1B (with 64 and 112 compounds detected in green and roasted beans respectively). These compounds included 57 VOCs, 21 amino acids, eight sugars and sugar alcohols, 25 organic acids, and eight fatty acids (Supplementary Table 2). All of the amino acids, sugars and sugar alcohols, and fatty acids quantified were detected in both green and roasted fractions.

3.2.1. Metabolites in green coffee beans

We identified and quantified a total of 64 metabolites in green coffee beans. This set included 12 VOCs, 16 organic acids, 8 sugars/sugar alcohols, 8 fatty acids, and 21 amino acids (Supplementary Table 2). To identify which metabolites might be potential markers for the presence of defective beans, we performed differential enrichment analysis comparing defective and non-defective fractions in the first batch of coffee beans. Hierarchal clustering of these metabolites across a heat map of relative metabolite abundance (Figure 2) revealed a cluster of compounds (Cluster I) enriched in defective seeds. This cluster was strongly dominated by a mix of amino acids, sugars/sugar alcohols, and organic acids. Cluster I includes lysine, phenylalanine, glucose, fructose, and pentitols (*t*-test, $p < 0.05$; Supplementary Table 2). The enrichment of these amino

acids in defective beans is in agreement with previous studies (Mazzafera, 1999). Asparagine and sucrose were the most abundant amino acid and sugar in green coffee, consistent with previously reported analyses (Mazzafera, 1999). The levels of these two metabolites did not differ significantly between our pools of defective and non-defective beans as was previously reported (Mazzafera, 1999), who observed an enrichment of asparagine and sucrose specifically in immature seeds. This discrepancy could be due to the low incidence of immature seeds in the defective fractions used in this study. The main organic acids present in green samples were citric and malic acids (Supplementary Table 2), which have been shown to contribute to the acidity of the coffee brew (Galli & Barbas, 2004; Illy & Viani, 2005). The levels of these two main metabolites were not significantly different between defective and non-defective beans. Although minor, gluconic and mucic acids were found to be significantly higher in defective samples ($p < 0.05$) (Figure 2) (Supplementary Table 2), but their contribution to overall cup quality remains to be determined. Interestingly, gluconic acid has been associated with repression of plant response to herbivory during regurgitant insect feeding, suggesting a possible origin for higher levels in defective beans (Musser et al., 2002). None of the VOCs or fatty acids examined in green coffee beans presented significant differences between defective and non-defective samples (Supplementary Table 2).

3.2.2. Metabolites in roasted coffee beans

We identified a total of 112 compounds, which we examined for patterns of differential enrichment in roasted beans prepared from defective and non-defective fractions. This set included 54 VOCs, 21 organic acids, and all of 8 sugars/sugar alcohols,

8 fatty acids, and 21 amino acids. After clustering the compounds across a heat map of relative compound abundance, we identified two distinct clusters including compounds enriched in defective (Cluster I) and non-defective (Cluster III) beans (Figure 3). In the analyses of roasted beans, Cluster I was dominated by volatile compounds such as butanal-3-methyl, furan, 2-(2-furanylmethyl)-5-methyl- (aka 5-methyl-2-furfurylfuran), pyridine, 2-methyl furan and 1-methylpyrrole, and sugar alcohols, including pentitols and inositol. Organic acids and sugars such as citric and malic acids, sucrose, and fructose dominated Cluster III.

VOCs, such as pyridines, furans, pyrroles, and pyrazines, dominated the defect enriched Cluster I. Butanal-3-methyl, responsible in part for a buttery flavor (Farah, 2012), was the most abundant VOC accumulating in defective samples. However, this volatile has not been previously associated with a particular defect. Pyridine, also significantly enriched in defective seeds ($p < 0.001$), is derived from trigonelline degradation during roasting and it gives an unpleasant aroma to coffee (Farah, 2012). Consistent with previous studies, our analyses also show that pyridine accumulates in a mix of defective beans (Toci & Farah, 2008). Levels of 5-methyl-2-furancarboxaldehyde (aka 5-methyl furfural) were in the same range as pyridine, and it gives coffee the almond spicy notes found is some foods (Table 1) (Petisca, Perez-Palacios, Farah, Pinho, & Ferreira, 2013). Finally, 2-methyl furan and 1-methyl-1-H-pyrrole were produced at comparable levels. Both compound classes have been previously shown to accumulate in defective beans (Mancha Agresti, Franca, Oliveira, & Augusti, 2008). Interestingly, 1-hexanol was not detected in defective green coffee in our experimental conditions, but was found in defective roasted samples. This compound was previously identified as a

possible marker for defective coffee and detected in three to ten-fold higher levels in control (non-defective) green samples (Toci & Farah, 2008), but not in roasted samples. These discrepancies may be explained by the different origins of the coffee and/or different roasting conditions, and they highlight that defective coffee markers may be limited to specific bean origins.

Upon roasting, sucrose and inositol were the main sugar and sugar alcohol in roasted coffee beans prepared from both non-defective and defective samples (Supplementary Table 2). While free sugars (fructose, glucose and sucrose) were significantly higher in non-defective beans ($p < 0.0001$ respectively) (Figure 3, Cluster III), the sugar alcohols (hexitols, inositol, pentitols and tetraols) were significantly higher in defective seeds ($p = 0.003$ and 0.005 for hexitol and inositol respectively and $p < 0.0001$ for pentitols and tetraols) (Figure 3, Cluster I). Overall, amino acid levels dropped to approximately 1% of green coffee levels upon roasting (Supplementary Table 2). Serine was the main amino acid present in roasted coffee beans, although no significant differences were found between defective and non-defective samples. Alanine, aspartate, threonine, and tryptophan levels were significantly higher in non-defective roasted beans than defective roasted seeds ($p = 0.008$, 0.039, 0.012, and 0.041 respectively; Figure 3, Cluster III; Supplementary Table 2). These results, together with the higher levels and numbers of VOCs measured in roasted defective beans, indicate that the Maillard and Strecker reactions occurred to a larger extent in defective beans.

For roasted beans, we found that several organic acids, such as citric, malic, and pyruvic acids, were enriched in non-defective samples (Fig 3a, Cluster III). While it is known that organic acids contribute in large part to the acidity of the coffee brew (Ginz,

Balzer, Bradbury, & Maier, 2000), in this study the differences in organic acid levels between defective and non-defective fractions was not reflected in the pH or titratable acidity of green or roasted coffee beans (Supplementary Table 1).

Linoleic and palmitic acids were the main fatty acids accumulating in green coffee beans (Supplementary Table 2). Fatty acids have been shown to be important in coffee bean quality (Farah, 2012), by being associated with coffee freshness and prevention of hydrolysis and oxidation. We found no significant differences between non-defective and defective fractions of green coffee beans in this study. In addition, fatty acid levels did not change substantially after roasting (Supplementary Table 2). These results are in accordance with a previous study in which fatty acid content was determined in solvent-extracted and screw-pressed coffee bean oil comparing non-defective and defective green coffee beans (Oliveira, Franca, Camargos, & Ferraz, 2008; Oliveira et al., 2006). In this study system, fatty acids appear to not be related to the observed coffee defects.

3.3. Determination of marker compounds

3.3.1 Marker identification

To further narrow down which of the metabolites accumulating in defective coffee could be utilized as biochemical markers, we analyzed the metabolic profiles for green and roasted beans aiming to define a core set of compounds that are consistently detected in the presence of defects. We conducted a PCA with the set of 64 compounds detected in green coffee for batch 1 to identify which of these analytes were driving the separation of defective and non-defective fractions (Figures 4 A). From the PCA

loadings in green bean fractions, a subset of 17 compounds separated the defective and non-defective fractions, accounting for 66.8% of the variation in the data set (54% and 12.8% on PC1 and PC2, respectively; Figures 4 A and B). This core-set of green coffee markers was then validated on the second batch of coffee beans, in which it separated defective and non-defective fractions with a similar power, accounting for 63.8% of the variation in the data set (39.1% and 24.7% on PC1 and PC2, respectively; Supplementary Figure 1).

From the 112 potential defective bean markers we identified for roasted coffee in batch 1, a core-set of 36 compounds separated defective and non-defective beans via PCA loadings, accounting for 91.4% of the variation in the data set (84.1% and 7.3% on PC1 and PC2, respectively; Figures 5 A and B). This set of roasted coffee markers was then validated on the second batch of coffee beans. One marker, ethylpyrazine, was removed due to its absence in the second batch of coffee, leaving 35 core compounds for analysis. These markers for coffee defects separated defective and non-defective fractions with a similar effectiveness, accounting for 79% of the variation in the data set (64.7% and 14.3 % on PC1 and PC2, respectively; Supplementary Figure 2). The presence of ethylpyrazine in one batch of defective coffee but not in another highlights the variability in defective coffee marker accumulation depending on coffee batch. It is hence important to ensure that the biochemical markers are present in two independent coffee batches.

3.3.2 Marker set validation

The utility of the aforementioned marker compounds for detecting defects in unsorted coffee batches was then assessed by examining their distribution in defective, non-defective, and unsorted green and roasted coffee fractions. The 17 core markers for green coffee were present in all three fractions (Figure 6 A), and 30 of the core roasted coffee markers were present among all three fractions (Figure 6 B). These markers provided the best targets for monitoring defective beans given their differential enrichment between defective and non-defective fractions, in addition to being detected at the extremes of defective bean content. Three of the five marker compounds identified in roasted coffee, 1H-pyrrole, 1-methyl-; furan, 2-(2-furanylmethyl)-5-methyl-; and furan, 2-methyl-, were uniquely enriched in defective roasted coffee beans only (Table 1). These volatile compounds are known to accumulate in defective beans (Mancha Agresti et al., 2008; Petisca et al., 2013), though their unique enrichment in this system may highlight their importance to the coffee defects in this study.

The detection of these markers in unsorted batches in this system would indicate a defect level above 7%, though their applicability to other cultivars or regions remains to be determined. Pyridine, enriched here in defective and unsorted roasted fractions (Table 1), has been associated with defective coffee beans and darkly roasted coffee due to the thermal degradation of trigonelline (Farah, 2012; Gökmen, 2015; Toci & Farah, 2008, 2014). Its presence in the defective and unsorted beans could be an indicator that the defective beans present in both of these fractions have altered roasting properties, allowing chemical reactions due to roasting, like the Millard and Strecker reactions, to occur to a further extent in defective beans. Uniformity in the degree of coffee roasting is vital in controlling the organoleptic properties of the brewed beverage, especially in

avoiding burning, or tipping, that can occur if beans are not uniform (Illy & Viani, 2005). Acrylic acid was present only in unsorted and non-defective fractions (Table 1), suggesting that it could be impacted similarly by the altered roasting properties of defective beans, resulting in its reduction below detectable limits in highly defective coffee, due to its increased conversion to acrylamide in the more highly roasted defective beans (Illy & Viani, 2005; Gökmen, 2015). The enrichment of pyridine and acrylic acid in our defective samples hints at the thermal origins of the coffee defects in our samples and suggest that they could serve as markers for roast degree.

4. Conclusion

Our multipronged approach to marker identification was successful in identifying a core set of compounds capable of discriminating between defective and non-defective coffee fractions in both green and roasted coffee. While a number of compounds were differentially enriched between fractions, the marker compounds identified in this study serve as a focused list that have been validated on independent coffee batches from Sumatra. Future studies should reveal the extent to which the markers identified in this study are generally useful regardless of cultivar and/or country of origin. Additionally, the approach to identifying marker compounds detailed here could be applied to identify those compounds associated with specific defects. Such studies would greatly contribute to the ability of processors to focus efforts on removing those defects with the greatest influence on perceived quality.

Acknowledgements

We are grateful to the Center for Applied Plant Sciences Targeted Metabolomics Laboratory (metabolomics.osu.edu) for access to the GC-MS and LC-MS/MS equipment funded by the Translational Plant Sciences Targeted Investment in Excellence (TIE). Finally, the authors would like to thank the other members of the Opportunity Team for Coffee Research at The Ohio State University, Joshua Blakeslee, Thaddeus Ezeji, and Thomas Mitchell, for their insights and contributions to this project.

Funding sources

This research was funded by the Center for Applied Plant Sciences. MJV salary was supported by the J. M. Smuckers Company.

Conflict of interest

The authors declare no conflict of interest in the preparation of this manuscript.

REFERENCES

Batista, L. R., & Chalfoun, S. M. (2014). Quality of Coffee Beans. In *Cocoa and Coffee fermentations* (pp. 477-508): CRC Press.

Cocuron, J. C., & Alonso, A. P. (2014). Liquid chromatography tandem mass spectrometry for measuring (1)(3)C-labeling in intermediates of the glycolysis and pentose phosphate pathway. *Methods Mol Biol, 1090*, 131-142.

Cocuron, J. C., Anderson, B., Boyd, A., & Alonso, A. P. (2014). Targeted metabolomics of Physaria fendleri, an industrial crop producing hydroxy fatty acids. *Plant Cell Physiol, 55*(3), 620-633.

Farah, A. (2012). Coffee constituents. In Y. Chu (Ed.), *Coffee: emerging health effects and disease prevention* (pp. 21-58): John Wiley & Sons, Blackwell Publishing Ltd.

Flament, I. (2002). *Coffee flavor chemistry*. Baffins Lane: John Wiley & Sons, Ltd.

Franca, A. S., Mendoca, J. C. F., & Oliveira, S. D. (2005). Composition of green and roasted coffees of different cup qualities. *LWT-Food Science and Technology, 38*, 709-715.

Franca, A. S., Oliveira, L. S., Mendoca, J. C. F., & Silva, X. A. (2005). Physical and chemical attributes of defective crude and roasted coffee beans. *Food Chemistry, 90*, 89-94.

Galli, V., & Barbas, C. (2004). Capillary electrophoresis for the analysis of short-chain organic acids in coffee. *Journal of Chromatography A, 1032*, 299-304.

George, S. E., Ramalakshmi, K., & Rao, L. J. M. (2008). A perception on health benefits of coffee. *Critical Reviews in Food Science and Nutrition, 48*, 464-486.

Ginz, M., Balzer, H. H., Bradbury, A. G. W., & Maier, H. G. (2000). Formation of aliphatic acids by carbohydrate degradation during roasting of coffee. *European Food Research and Technology, 211*(6), 404-410.

Gökmen, V. (2015). *Acrylamide in food. Analysis, content an potential health effects* (1 ed.). Boston: Elsevier.

Illy, A., & Viani, R. (2005). *Espresso coffee: the science of quality* (2nd ed.). Boston: Elsevier.

Kosalos, J., Stephen, R., & Diaz, S. (2013). *Arabica green coffee defect handbook*: Specialty Coffee Association of America.

Koubaa, M., Cocuron, J. C., Thomasset, B., & Alonso, A. P. (2013). Highlighting the tricarboxylic acid cycle: liquid and gas chromatography-mass spectrometry analyses of (13)C-labeled organic acids. *Anal Biochem, 436*(2), 151-159.

Mancha Agresti, P. D. C., Franca, A. S., Oliveira, L. S., & Augusti, R. (2008). Discrimination between defective and non-defective Brazilian coffee beans by their volatile profile. *Food Chemistry, 106*, 787-796.

Mazzafera, P. (1999). Chemical composition of defective coffee beans. *Food Chemistry, 64*, 547-554.

Mendoca, J. F. C., Franca, A. S., Oliveira, L. S., & Nunes, M. (2008). Chemical characterization of non-defective and defective green Arabica and Robusta coffees by Electrospray Ionization-Mass Spectrometry. *Food Chemistry, 111*, 490-497.

Musser, R. O., Hum-Musser, S. M., Eichenseer, H., Peiffer, M., Ervin, G., Murphy, J. B., et al. (2002). Herbivory: caterpillar saliva beats plant defences. *Nature, 416*(6881), 599-600.

Oliveira, L. S., Franca, A. S., Camargos, R. R., & Ferraz, V. P. (2008). Coffee oil as a potential feedstock for biodiesel production. *Bioresource Technology, 99*, 3244-3250.

Oliveira, L. S., Franca, A. S., Mendoca, J. C. F., & Barros-Junior, M. C. (2006). Proximate composition and fatty acids profile of green and roasted defective coffee beans. *LWT-Food Science and Technology, 39*, 235-239.

Petisca, C., Perez-Palacios, T., Farah, A., Pinho, O., & Ferreira, I. M. P. L. V. O. (2013). Furans and other volatile compounds in ground roasted and espresso coffee using headspace solid-phase microextraction: Effect of roasting speed. *Food and Bioproducts Processing, 91*, 233-241.

Ramalakshmi, K., Kubra, I. R., & Rao, L. J. M. (2007). Physicochemical Characteristics of Green Coffee: Comparison of graded and defective beans. *Journal of Food Science, 72*(5), 333-337.

Toci, A. T., & Farah, A. (2008). Volatile compounds as potential defective coffee beans' markers. *Food Chemistry, 108*, 1133-1141.

Toci, A. T., & Farah, A. (2014). Volatile fingerprint of Brazilian defective coffee seeds: corroboration of potential marker compounds and identification of new low quality indicators. *Food Chemistry, 153*, 298-314.

van Boekel, M. A. J. S. (2006). Formation of flavour compounds in the Maillard reaction. *Biotechnology Advances, 24*, 230-233.

Xia, J., Mandal, R., Silenikov, I., Broadhurst, D., & Wishart, D. S. (2012). MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis. *Nucleic Acid Research, 40*(1), 127-133.

Xia, J., Psychogios, N., Young, N., & Wishart, D. S. (2009). MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research, 37*, 652-660.

Xia, J., Silenikov, I., Han, B., & Wishart, D. S. (2015). MetaboAnalyst 3.0 - making metabolomics more meaningful. *Nucleic Acids Research, 43*, 251-257.

TABLES

Table 1. Biochemical markers with their organoleptic properties. A subset of the compounds described have known impacts on the organoleptic properties of the coffee brew and associations with coffee defects.

Uniquely Enriched Compound	Detected in	Assay	Class	Organoleptic properties
Acrylic acid	Non-defective/ Unsorted Roast	OA	Carboxylic acids	Described in green and roasted coffee, unknown contribution (Flament, 2002)
Pyridine	Defective/ Unsorted Roast	HS	Pyridines	Pungent and malodorous at high concentrations; warm, burnt, smoky at low concentrations (Farah, 2012; Flament, 2002; Gökmen, 2015; Toci & Farah, 2008)
1H-Pyrrole, 1-methyl-	Defective Roast Only	HS	Pyrrole	Odor smoky or tarry, flavor beany or metallic (Flament, 2002; Gökmen, 2015; Toci & Farah, 2008)
Furan, 2-(2-furanylmethyl)-5-methyl-	Defective Roast Only	HS	Furan	Flavor earthy or fungal (Flament, 2002; Toci & Farah, 2008)
Furan, 2-methyl-	Defective Roast Only	HS	Furan	Odor ethereal or sickly (Flament, 2002; Petisca et al., 2013)

FIGURE CAPTIONS

Figure 1. Sampling and biochemical analysis work flow. (A) Coffee seeds were sorted manually from an unsorted pool and split into defective and non-defective, generating three groups for analyses. Defects found in the unsorted pool included ears, black beans, insect damage, discoloration, necrosis and mechanical damage. Each group was subjected to roasting and subsequent biochemical profiling. **(B)** Green and roasted seeds were analyzed for extracted and analyzed for amino acids (AA), sugars and sugar alcohols (S/SA), organic acids (OA), headspace volatiles (HS), and fatty acid methyl esters (FAMEs). Two screening methods were used to identify potential markers for defective green and roasted coffee. These potential markers were validated on a separate batch of coffee beans. Markers present in both sets were identified as defective coffee markers.

Figure 2. **Compound enrichment in defective and non-defective green coffee beans.** Ward's hierarchical clustering of detected compounds imposed over a heat map showing the relative abundance of each compound for defective (n=4, D1-D4) and non-defective (n=4, ND1-ND4) green coffee samples. Cluster I includes compounds enriched in defective coffee seeds. Enrichment in the candidate markers was tested using a 1-sample *t*-test.

Figure 3. Compound enrichment in defective and non-defective roasted coffee beans. Ward's hierarchical clustering of detected compounds a) imposed over a heat map showing the relative abundance of each compound for defective (n=4, D1-D4) and non-defective (n=4, ND1-ND4) roasted coffee samples. Compounds enriched in defective

coffee seeds are highlighted in Cluster I, and compounds enriched in non-defective coffee seeds are highlighted in Cluster III. Cluster II is a group of analytes whose levels were not distinct between defective and non-defective roasted coffee beans.

Figure 4. **Biochemical marker identification in defective green coffee beans.** Serial reductive PCA identified a subset of metabolites maximizing variability in the dataset, which identified defective and non-defective fractions (D for defective, ND for non-defective). Shaded regions in PCA plots represent 95% confidence intervals for each sample type. Enrichment in the candidate markers was also tested using a 1-sample t-test. Candidate marker compounds followed by (*) were significantly enriched at $p < 0.05$.

Figure 5. Biochemical marker identification in defective roasted coffee beans. Serial reductive PCA identified a subset of metabolites maximizing variability in the dataset, which identified defective and non-defective fractions (D for defective, ND for non-defective). Shaded regions in PCA plots represent 95% confidence intervals for each sample type. Enrichment in the candidate markers was also tested using a 1-sample t-test. As similarly done for green bean samples, candidate marker compounds followed by (*) were significantly enriched at $p < 0.05$.

Figure 6. Validation of biochemical markers. The 17 green coffee a) defect markers were detected in all three coffee fractions. However, in the roasted seeds b), five compounds were uniquely enriched among separate fractions.

SUPPORTING INFORMATION

Supplementary Figure 1. Principal component score plot showing the separation of defective samples (n=4; Pink shading, D1- D4) and non-defective samples (n=4; Green shading, ND1-ND4) samples using the core marker set identified for green coffee. Shaded regions represent 95% confidence intervals for each sample type.

Supplementary Figure 2. Principal component score plot showing the separation of defective samples (n=4; Pink shading, D1-D4) and non-defective samples (n=4, Green shading, ND1-ND4) samples using the core marker set identified for roasted coffee. Shaded regions represent 95% confidence intervals for each sample type.

Supplementary Table 1. Physico-chemical properties of green and roasted coffee beans.

Supplementary Table 2. Quantification data for each compound family studied.

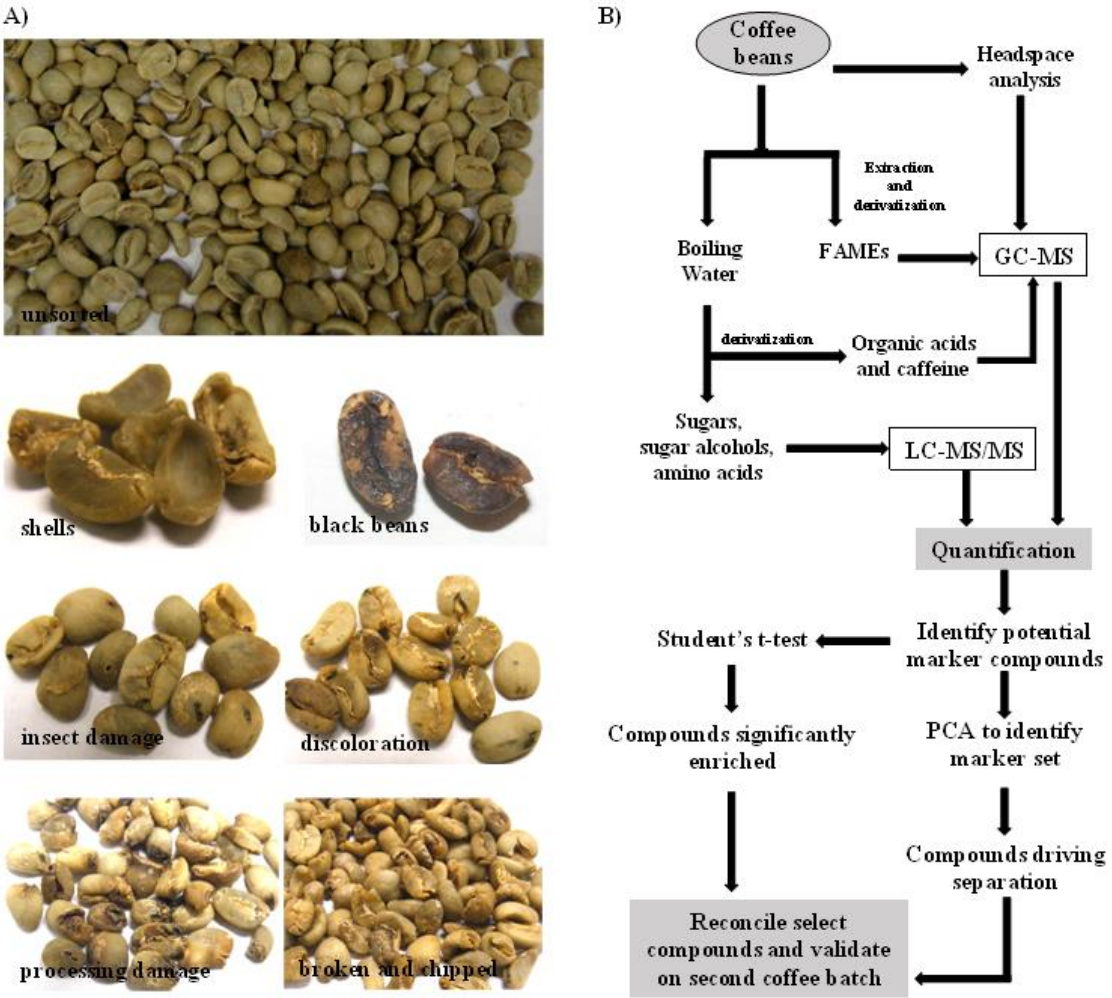


Fig. 1

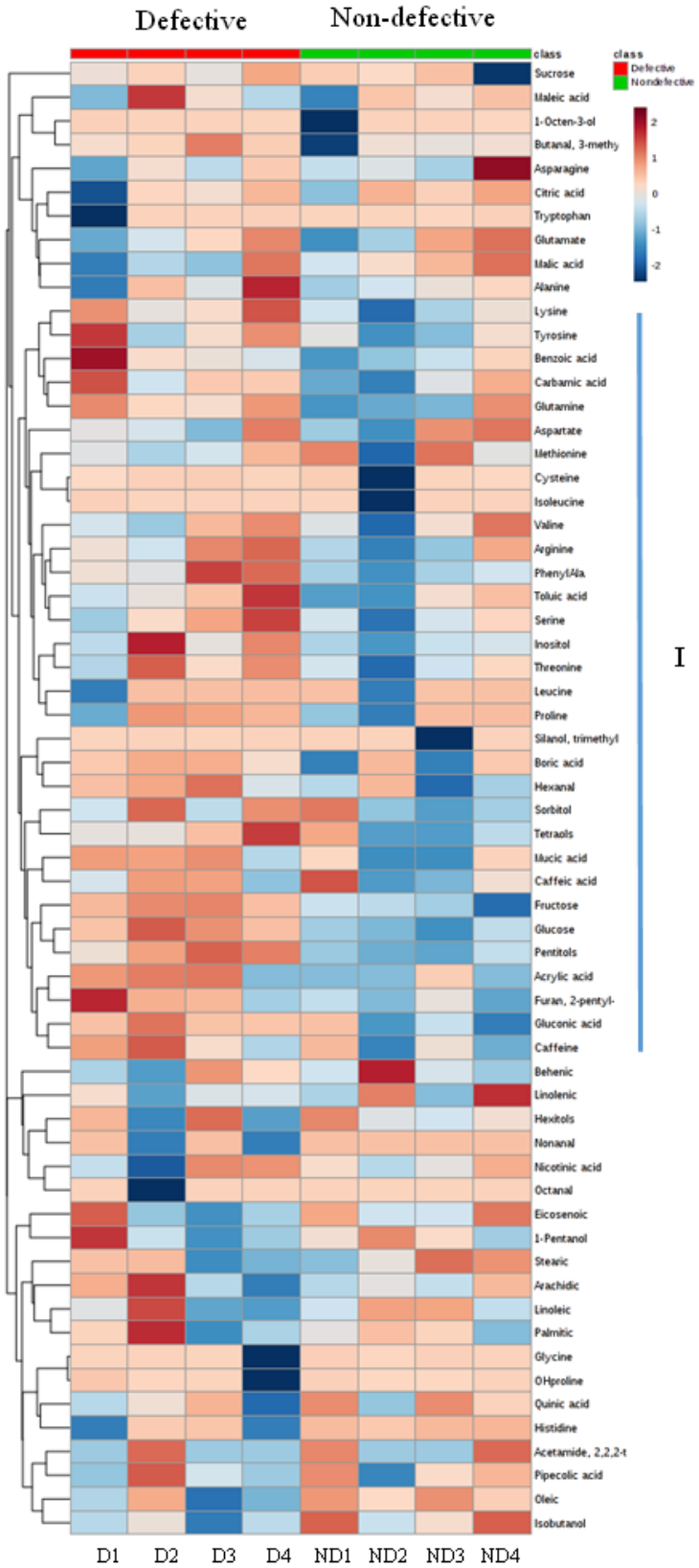


Fig. 2

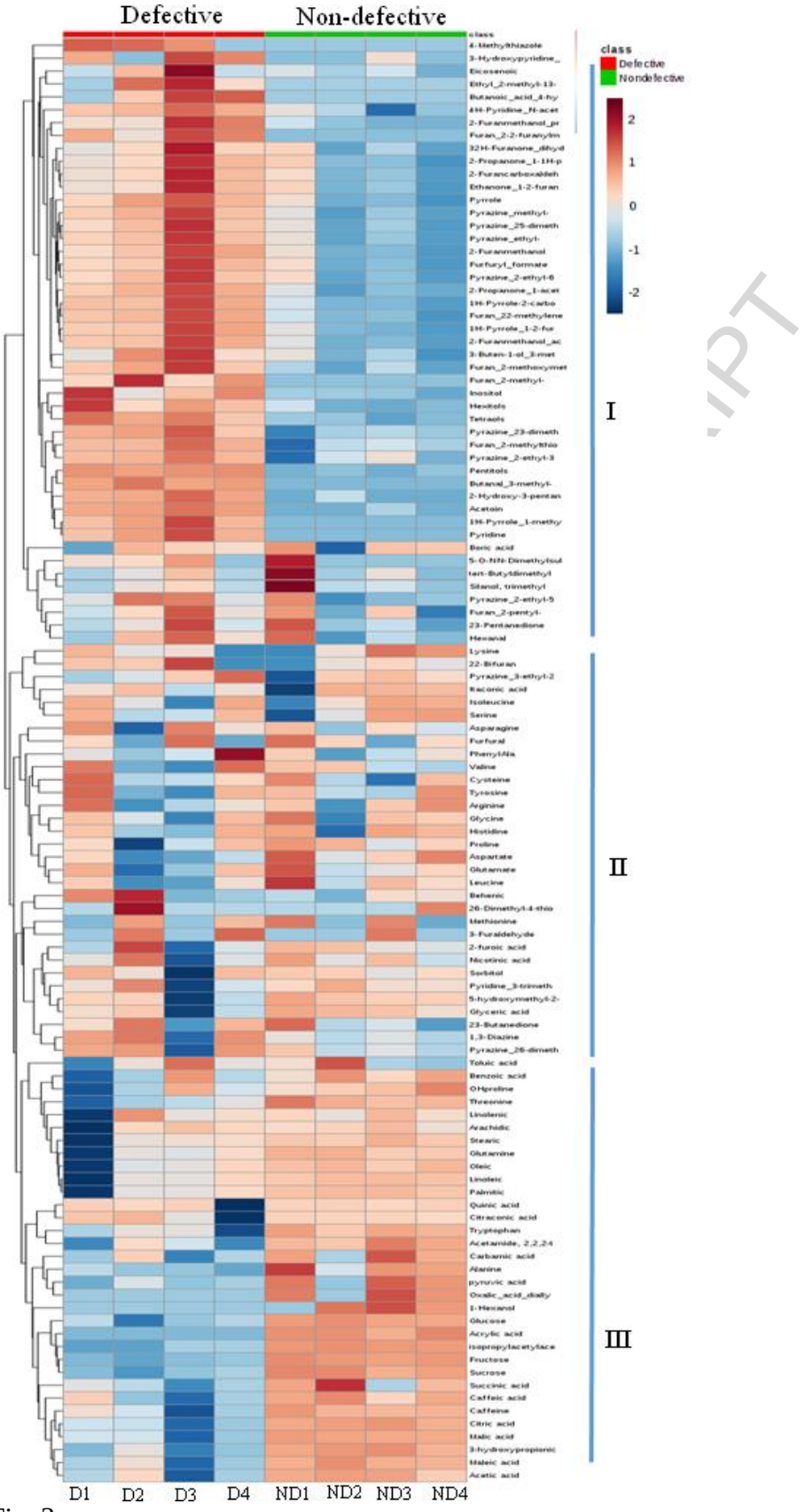
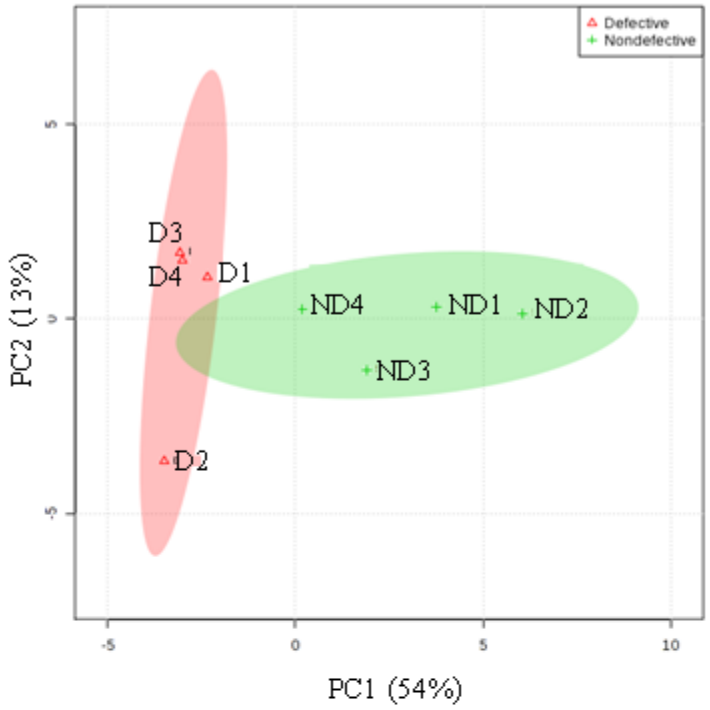


Fig. 3

A)

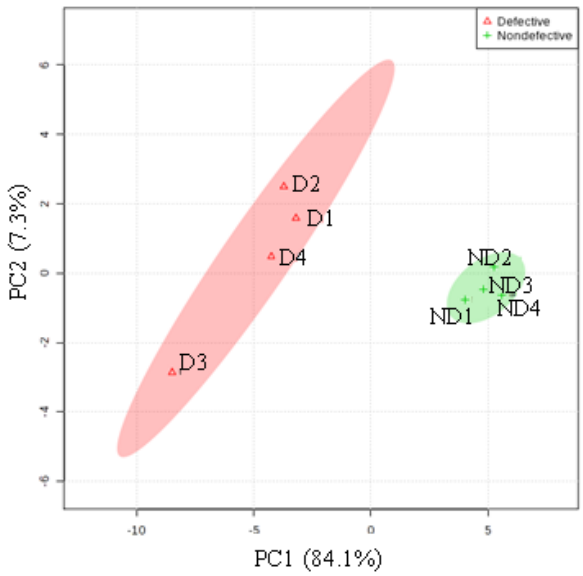


B)

Potential Marker Compounds	Compound Class
Arginine	AA
Glutamine	AA
Isoleucine	AA
Lysine*	AA
Phenylalanine*	AA
Threonine	AA
Tyrosine	AA
Behenic	FA
Oleic acid	FA
Nonanal	HS
Acrylic acid	OA
Benzeneacetic acid	OA
Mucic acid	OA
Fructose*	S/SA
Glucose*	S/SA
Inositol	S/SA
Pentitols *	S/SA

Fig. 4

A)



B)

Potential Marker Compound	Compound Class
1H-Pyrrole_1-(2-furanylmethyl) *	HS
1H-Pyrrole_1-methyl *	HS
1H-Pyrrole-2-carboxaldehyde_1-methyl *	HS
2-Furanmethanol_acetate *	HS
2-Hydroxy-3-pentanone *	HS
2-Propanone_1-(acetyloxy)- *	HS
3-Hydroxypropionic acid *	OA
2-Furoic acid	OA
Acetic acid	OA
Acetoin *	HS
Acrylic acid *	OA
Alanine *	AA
Aspartate	AA
Butanal_3-methyl *	HS
Citric acid *	OA
Caffeine	Caff
Caffeic acid	OA
Fructose *	S/SA
Furan_2-(2-furanylmethyl)-5-methyl- *	HS
Furan_2-(methoxymethyl)- *	HS
Furan_2-methyl- *	HS
Furan_2-methylenebis- *	HS
Glucose *	S/SA
Glyceric acid	OA
Hexitols *	S/SA
Inositol *	S/SA
Maleic acid	OA
Malic acid *	OA
Pentitols *	S/SA
Pyrazine, 2,5-dimethyl-	HS
Pyrazine, ethyl-	HS
Pyrazine_methyl- *	HS
Pyridine *	HS
Pyrrole *	HS
Sucrose *	S/SA
Tetraols *	S/SA

Fig. 5

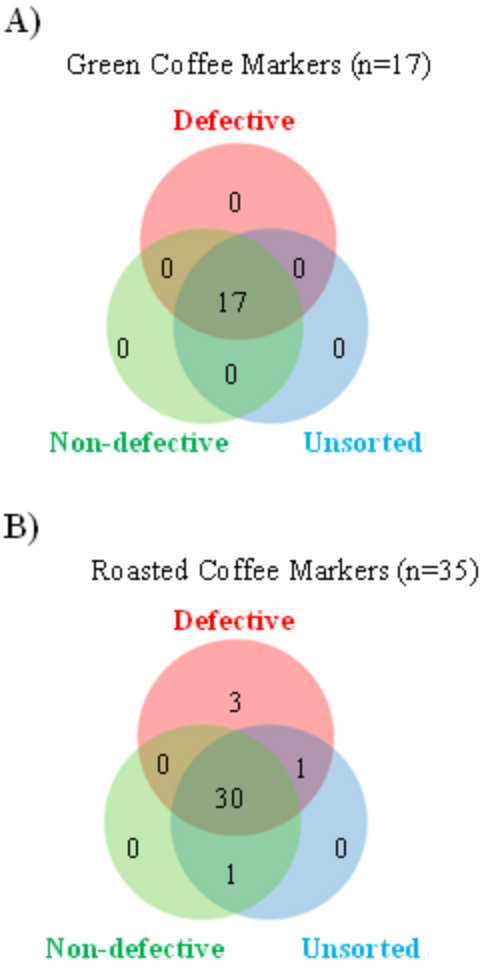


Fig. 6

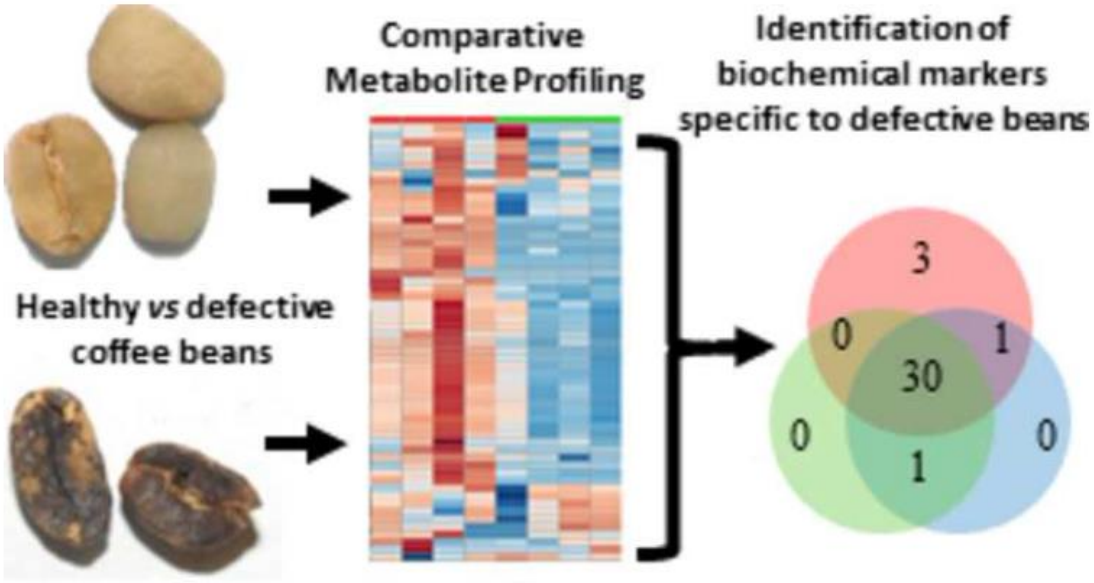
HIGHLIGHTS

• Biochemical markers for defective coffee seeds were identified.

• Defective green seeds were enriched in reducing sugars and free amino acids.

• Defective roasted beans were enriched in volatiles and sugar alcohols.

• Volatiles unique to defective roasted coffee seeds are known to alter coffee flavor.



Graphical abstract