FOODCHEM-D-18-04215 Manuscript revision

Title:

Untargeted Mass Spectrometry-Based Metabolomics Approach Unveils Molecular Changes in Raw and Processed Foods and Beverages

Authors:

Julia M. Gauglitz[a,b,c], Christine M. Aceves[a], Alexander A. Aksenov[a,b], Gajender Aleti[d], Jehad Almaliti[e,f], Amina Bouslimani[a,b], Elizabeth A. Brown, Anaamika Campeau[b,g], Andrés Mauricio Caraballo-Rodríguez[a,b], Rama Chaar[b], Ricardo R. da Silva[a,b], Alyssa M. Demko[e], Francesca Di Ottavio[h], Emmanuel Elijah[a,b], Madeleine Ernst[a,b], L. Paige Ferguson[b], Xavier Holmes[a,b], Alan K. Jarmusch[a,b], Lingjing Jiang[j], Kyo Bin Kang[a,b], Irina Koester[e], Brian Kwan[j], Jie Li[e], Yueying Li[e], Alexey V. Melnik[a,b,c], Carlos Molina-Santiago[k], Bohan Ni[e], Aaron L. Oom[l], Morgan W. Panitchpakdi[a,b], Daniel Petras[a,b,c,5], Robert Quinn[a,b], Nicole Sikora[b], Katharina Spengler[a], Bahar Teke[e], Anupriya Tripathi[a,b], Sabah Ul-Hasan[m], Justin J.J. van der Hooft[a,b,l], Fernando Vargas[a,b,n], Alison Vrbanac[o], Anthony Q. Vu[o], Steven C Wang[b,n], Kelly Weldon[a,b,c], Kayla Wilson[e], Jacob M. Wozniak[b,g], Michael Yoon[b], Nuno Bandeira[b,p], Pieter C. Dorrestein[a,b,q*]

Author Affiliations:

a. Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego.
b. Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego
c. Center for Microbiome Innovation, University of California, San Diego.
d. Mammalian Genomics, J. Craig Venter Institute, San Diego.
e. Scripps Institution of Oceanography, University of California, San Diego
f. Department of Pharmaceutical Sciences, Faculty of Pharmacy, The University of Jordan, Amman, 11942, Jordan
g. Department of Pharmacology, University of California, San Diego.
h. University of Teramo, Faculty of Bioscience, Te, Italy
i. Bioinformatics Group, Wageningen University, Wageningen, The Netherlands
j. Department of Family Medicine and Public Health, University of California, San Diego.
k. Departamento de Microbiología, Instituto de Hortofruticultura Subtropical y Mediterránea “La Mayora”, Universidad de Málaga, Bulevar Louis Pasteur 31 (Campus Universitario de Teatinos), 29071 Málaga, Spain.
l. Department of Medicine, University of California, San Diego

m. School of Natural Sciences, University of California Merced, Merced, CA, 95343, USA
n. Division of Biological Sciences, University of California at San Diego, La Jolla, CA
o. Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA, USA
p. Department of Computer Science and Engineering, University of California, San Diego.
q. Departments of Pharmacology and Pediatrics, University of California, San Diego.

Abstract:

In our daily lives, we consume foods that have been transported, stored, prepared, cooked, or otherwise processed by ourselves or others. Food storage and preparation have drastic effects on the chemical composition of foods. Untargeted mass spectrometry analysis of food samples has the potential to increase our chemical understanding of these processes by detecting a broad spectrum of chemicals. We performed a time-based analysis of the chemical changes in foods during common preparations, such as fermentation, brewing, and ripening, using untargeted mass spectrometry and molecular networking. The data analysis workflow presented implements an approach to study changes in food chemistry that can reveal global alterations in chemical profiles, identify changes in abundance, as well as identify specific chemicals and their transformation products. The data generated in this study are publicly available, enabling the replication and re-analysis of these data in isolation, and serve as a baseline dataset for future investigations.

Keywords: Untargeted mass spectrometry, Metabolomics, Molecular networking, LC-MS/MS, Food, Fermentation, Tea; Yogurt

1. Introduction:

We consume a variety of foods and beverages during any given day, such as fruits, vegetables, dairy products and meats. The chemical composition of these foods is influenced by factors such as the source, processing method, storage or other handling before consumption, which has been a central focus of the food science field. However, new measurements and data analysis methods can help expand and clarify our understanding of the molecular composition of foods. Within the food science field, there is significant interest and awareness of dietary habits of human populations (Lewis et al., 2009; Schulze et al., 2015-2020 Dietary Guidelines for Americans) and the nutritional composition of food (Thirumdas et al., 2018). Resultant findings can generate policies and nutritional recommendations with the end goal of improving public health (Berger et al, 2019).

Mass spectrometry (MS) is an analytical tool that detects ionized molecules and can be used for identification and quantification. The majority of food MS studies employ targeted analysis of a set of predefined compounds via gas and liquid chromatography-mass spectrometry (GC-MS and LC-MS). Many food MS studies monitor chemicals that are harmful when consumed, but the chemical composition of food and its impact on health is not limited to these chemicals (Giorio et al., 2017, Scalbert et al., 2014). Furthermore, the utilization of MS is significant and expected to grow (Yoshimura et al., 2016) in areas such as food monitoring during processing (Marshall et al., 2017), due in part to the cost per data volume of MS having decreased by two orders of magnitude over the past 15 years, and the prediction that it is expected to continue to decrease, presenting MS as a feasible method for large datasets (Aksenov et al., 2017). We present an untargeted approach using liquid chromatography-tandem mass spectrometry (LC-MS/MS) to illustrate the effects of storage

and processing on different food types and, for the first time, pair this methodology with emerging MS-based computational analysis approaches, such as mass spectral molecular networking to assess changes based on processing.

Mass spectral molecular networking enables a broad overview of molecular information that can be inferred from MS/MS data (Watrous et al., 2012). In molecular networking, all identical MS/MS spectra are merged giving a list of unique MS/MS spectra (Watrous et al., 2012). These are then subjected to spectral alignment allowing for spectral matching with offsets based on the precursor mass differences. Molecules generating similar MS/MS spectra are clustered due to similarities in their fragmentation patterns and are referred to as molecular families. A molecular family is a set of MS/MS spectra that are structurally related (Nguyen et al., 2013). In addition, the MS/MS spectra are putatively annotated against reference spectra within the Global Natural Products Social Molecular Networking (GNPS) platform (Yang et al., 2013, Wang et al., 2016). Matches against the reference libraries constitute level 2 or 3 annotations according to the 2007 metabolomics standards initiative (Sumner et al., 2007). The reference libraries that can be searched, as their spectra are publicly available or available for purchase, include: NIST17, Massbank Europe and North America, ReSpect, CASMI, EMBL metabolomics library, HMDB, and GNPS contributed MS/MS spectra (Wang M et al., 2016, Aksenov et al., 2017, Blaženović et al., 2018). The resulting molecular networks visualize chemical relationships of compounds and provide a powerful tool for in-depth interpretation of chemical transformations. One example of this is the use of molecular networking to help characterize a large number of triterpene saponins in Siberian ginseng (Ge et al., 2017)

105 We hypothesize that the untargeted metabolomics approach presented provides
106 information within a gap between targeted molecular analysis and elemental and
107 macronutrient analysis used in food chemistry. We demonstrate the utility of molecular
108 networking and other analysis tools, such as multivariate statistics, in analyzing untargeted
109 metabolomics data collected to assess the chemical impact of food handling techniques. The
110 potential for untargeted mass spectrometry to augment the knowledge of chemical processes
111 was assessed using the following well studied processes: 1) the impact of starter cultures on
112 the fermentation of yogurts, 2) the effects of brewing time on tea, 3) the effect of roasting
113 coffee on molecular composition, 4) how improper meat storage affects changes in chemistry,
114 and 5) how the molecular composition of tomato changes, depending on whether it was
115 ripened on or off the vine or the cultivar selected. All of these scenarios represent typical
116 processing situations that might occur in commonly consumed foods.

117 **2. Materials and Methods:**

118 We provide a general overview of the materials used and methodology for the five food types.
119 Details are provided where these are shared between most sample types - experimental
120 details for each food type as well as brand information can be found in the Supplementary
121 Information S1.1 - S2.2 and Table S1. 702 samples were considered in the final analysis.
122

123 *2.1 Sample collection*

124 Milk, yogurt, tea leaves, brewed tea, coffee beans, brewed coffee, turkey, beef, and tomato
125 were all sampled in duplicate; one replicate was extracted for analysis (see Supplemental
126 Material Table S1.) and the other was archived for future uses. Unique barcode numbers

were assigned to each sample. Liquid samples (defined here as milk, brewed tea, and brewed coffee) were collected into two identical empty 2 mL round bottom tubes (Qiagen, Hilden, Germany). All other sample types were collected in 2 mL round bottom tubes pre-filled with 1.0 mL room temperature ethanol-water (95:5 *v/v)*, (ethyl alcohol, pure, 200 proof (Sigma-Aldrich, Saint Louis, MO, USA)) and deionized Water (Invitrogen UltraPure™, Grand Island, NY, USA)). Duplicate samples were collected in empty 2 mL round bottom tubes and archived. Sample tubes were weighed before and after sample collection, unless otherwise noted in the metadata. All samples were stored at -80°C until downstream sample preparation for MS-based metabolomics. **Figure 1** highlights representative examples of images associated with the five food types that were sampled.

Pasteurized whole milk (Horizon Organic Vitamin D Milk; Broomfield, CO, USA) and three yogurts (Oikos Plain Greek Nonfat Yogurt (Dannon, Horsham, PA, USA), Voskos Plain Greek Yogurt (Sun Valley Dairy, Sun Valley, CA, USA) and Kroger Plain Nonfat Greek Yogurt (Kroger, Cincinnati, OH, USA)) were sampled in biological triplicates and used to culture three separate batches of home-fermented yogurt, which were sampled over 6 days for a total of 126 samples (see section S1.1 for culturing and sampling description). The yogurts from Oikos and Voskos contained the same live active cultures (*S. thermophilus; L. bulgaricus; L. acidophilus; Bifidus; L. casei*), whereas Kroger contained *L. acidophilus, B. bifidum, and L. casei.*

Twelve teas representing six varieties of tea leaves (Oolong, white, black, green, pu'er and matcha green) were purchased (see supplemental for detailed brand information) and sampled in biological triplicate before brewing, 10 water blanks, and at 0.5 min, 1 min, 4 min and 240 min after addition of hot water, giving a total of 185 brewed tea samples.

38 unique types of coffee were purchased, representing different roasts, brands, and origins. Coffee beans and brewed coffee were sampled in biological duplicates. There was a total of 152 samples.

There was a total of 119 meat samples; two types of turkey (certified organic and conventional) as well as two types of beef (certified organic and conventional) were sampled in biological triplicates over a 5-day time course to investigate meat spoilage. Each meat product was sampled into two petri dishes: one sample was spiked with tetracycline (final concentration of 300 ppb residual tetracycline) while the other was treated with the vehicle (i.e. 70% EtOH). Although tetracycline is used commonly as a growth promoter for livestock in some countries, here, it was added to see the effects of this antibiotic on a 5-day food spoilage test (Granados-Chinchilla & Rodríguez, 2017).

120 tomato samples were sampled from 8 different types of tomatoes (3 brands of conventional cherry tomato, 1 organic cherry tomato, 1 home-grown cherry tomato, 1 roma tomato from a San Diego [CA, USA] farmers' market, purchased canned tomatoes and sundried tomatoes) as well as a 5-day ripening time course of organic cherry tomatoes (see Supplemental for detailed brand information). The private garden-grown tomatoes were naturally grown, ripened on the vine and were not treated with any pesticides/herbicides. The farmers' market tomatoes were also indicated as not treated with pesticides/herbicides. We investigated the effect of origin and storage time (at room temperature) on the molecular composition.

Samples were collected according to detailed procedures outlined in the Supplementary Information (S1.1 and Table S1.) and depicted in **Figure 1**.

2.2 Metadata

Metadata were entered manually for all samples. Images were used to capture key sample information including unique barcode IDs, packaging information and time of sample collection. Metadata consisted of 142 different descriptive categories including but not limited to: ingredients, packaging type, location of food production, location of sample collection, store and brand names, UPC codes, NDB numbers and descriptions, cheese and dairy types, fermented and non-fermented food, botanical definitions and genus names of plant samples, conventional vs organically produced, type of animal meat, and presence of common allergens and additives. Sample information entries were standardized using a metadata dictionary that explained the types of information needed for each category as well as the correct formatting. The metadata spreadsheet and dictionary are publicly available (see Data and Code Availability in Appendix).

2.3 Sample processing

All samples were extracted in ethanol, centrifuged, dried by centrifugal evaporation and resuspended in 50% MeOH / 50% Water (Optima LC-MS grade; Fisher Scientific, Fair Lawn, NJ, USA) containing 2µM sulfadimethoxine (Analytical Standard, Sigma-Aldrich), as an injection control. Detailed sample processing information can be found in the Supplemental **Section S1.2**. 5 µL of resuspended extract was injected for LC-MS/MS analysis. Untargeted metabolomics was carried out using an ultra-high-performance liquid chromatography system (UltiMate 3000, Thermo Scientific, Waltham, MA) coupled to a Maxis Q-TOF (Bruker Daltonics, Bremen, Germany) mass spectrometer with a Kinetex C18 column (Phenomenex Torrance, CA, USA). Data were collected using a data dependent acquisition method outlined

196 in the Supplementary Information (**Section S2.1**). Electrospray ionization in positive mode was
197 used. Data were assessed for quality as described in **Section S2.2** prior to data analysis.
198
199 *2.4 Data analyses*
200 2.4.1 Molecular networking and small molecule annotations
201 GNPS molecular networking parameters were set to a minimum requirement of 4 ions to match
202 and a cosine score of >0.7 (https://gnps.ucsd.edu). Parent mass tolerance was 0.1 Da and MS/MS
203 was set to 0.1 Da (these parameters were used as many reference type spectra are low resolution).
204 The library search was performed with min match peaks of 4 and a cosine >0.7. Due to the different
205 small molecule compositions for each food, the annotations of all individual food analyses were
206 impacted differently, as recently shown with Passatutto, a false discovery rate (FDR) estimator
207 (Scheubert et al., 2017). Passatutto was used to estimate FDR for the annotations with our settings for
208 each of the five sub-analyses. Passatutto uses a decoy database created using fragmentation trees
209 and rebranching of fragments to estimate the FDR. With these analysis parameters, the estimated
210 FDR of annotations, based on spectral matching, at level 3 for the milk to yogurt was 0.5%, 0.2% for
211 tea, 0.09% for coffee, 1.5% for meat, and 4.8% for tomato.
212 For the milk to yogurt analysis, 126 samples resulted in 78,203 MS/MS spectra, of
213 which 63,241 passed the minimal requirement of four ions and minimum of two identical
214 spectra (**Supplementary Figure 6**). After clustering identical spectra 4,142 nodes remained.
215 147 of the nodes had spectral matches against the libraries searched (3.5% annotation rate).
216 185 tea samples resulted in 50,547 MS/MS spectra, 44,505 of which passed filtering
217 (**Supplementary Figure 10-11**). After merging identical spectra, 1,834 unique MS/MS
218 spectra comprised a molecular network with 207 annotations (11.2% annotation rate). 146
219 coffee samples resulted in a total of 50,929 MS/MS spectra. After filtering, 42,752 MS/MS

spectra remained, which condensed to 1,460 unique spectra in **Supplementary Figure 7**. Of the 1,460 unique spectra, 72 had spectral matches to the reference libraries within a cosine of 0.7 (4.9% annotation rate). The meat analysis included 119 samples, resulting in 72,083 MS/MS spectra, 54,663 of which passed the filtering step (**Supplementary Figure 8-9**). Merging all identical spectra resulted in 5,035 unique spectra of which 313 were annotated (6.2% annotation rate). MS of the 120 tomato samples resulted in 71,430 MS/MS spectra, 62,263 passed the filtering for a minimum of 4 ions and a minimum of two identical MS/MS spectra in the dataset, which condensed to 2,611 unique spectra that are presented as nodes (**Supplementary Figure 5**). 212 of the nodes were putatively annotated using the GNPS libraries (8.1% annotation rate). All annotations are level 2 or 3 according to the 2007 metabolomics standards initiative (Sumner et al., 2007).

2.4.2 Feature finding using mzMINE

MS^1 feature detection was performed using mzMINE2 (http://mzmine.github.io/). Outputs of the feature matrix report area-under-the-curve. Parameters used for feature finding can be found in Supplemental Materials (**Section S2.3**). Samples that did not contain the internal standard, sulfadimethoxine, were re-injected. MS/MS belonging to the internal standard sulfadimethoxine were observed in all data included in the analysis; this feature was removed from the MS^1 feature table prior to normalization by sample for downstream statistical analyses.

2.4.3 Multivariate statistical analysis and visualization

We used principal coordinates analysis (PCoA) to observe broad molecular patterns and trends within the data. PCoA takes a dissimilarity matrix as input and aims to produce a low-dimensional graphical representation of data, such that samples closer together have smaller dissimilarity values than those further apart. PCoA plots are a beta diversity metric (diversity between samples) and consist of orthogonal axes where each axis (PC1, PC2, PC3) captures a percentage of the total variance. For PCoA, signal intensities of the MS^1 features were normalized with Probabilistic Quotient Normalization (PQN) (Ejigu et al., 2013). PCoAs were calculated with the Canberra dissimilarity metric using QIIME (Caporaso et al., 2010) and visualized in EMPeror (Vázquez-Baeza et al., 2013).

Heatmaps were created from the filtered and preprocessed MS^1 feature tables, comprising both overall features as well as only features with a GNPS library hit. The Jupyter notebooks (R and python) used to create the heatmaps and perform the statistical analyses are publicly available at https://github.com/DorresteinLaboratory/supplementary-MolecularChangesInFood.

3. Results & Discussion

Untargeted MS revealed molecular differences between food types as well as within a food category due to variations in source and the time-based processing methods of fermentation, brewing, roasting, spoilage, and ripening. A combination of molecular networking, based on MS/MS spectra, multivariate and univariate statistical analysis of MS^1 features, and data visualization with principal coordinate analysis plots and heatmaps augmented current

chemical knowledge of these processes, and exemplified molecular differences on a global scale and individually for each food type.

3.1 A beta diversity analysis of food types and their processing

Visualization of the complex beta diversity matrix of our MS^1 data, visualized using PCoA plots, showed clear separation by sample type, which was expected (**Figure 2 and movie S1**). Yogurt and milk samples formed distinct groups (blue). Tomato and meat samples formed tight groups, orange and red, respectively, while the tea had two groups representing the solid (tea leaves) and brewed samples (green). Coffee had three groups corresponding to extracts of whole or ground beans, depending on the variety, and brewed coffee. Sample groups that had tighter clustering were more chemically similar, regardless of sample collection and processing. Because dissimilarities within samples were smaller than the dissimilarities between food types, each sample type was processed separately to maximize separations in PCoA space within a single food type.

PCoA analysis of the yogurt and milk samples showed separation based on brand, despite the fact that they contained similar live active cultures and ingredients **(Figure 3a)**. The home fermentation time courses of milk inoculated with different yogurts, as starter culture, are displayed in **Figure 3b and Figure S1a-S1b**. The fermentation process and associated molecular changes were visualized by all three home ferments becoming more yogurt-like, as illustrated by increases in distance between the Kroger yogurt and the starting milk in PCoA space, corresponding to transition of the home ferment (milk + starter yogurt) through time, becoming more similar to the original starter culture **(Figure 3b)**. Voskos contained Grade A pasteurized milk and cream, in addition to nonfat milk found in the Oikos,

286 possibly contributing to differences between these yogurts and the corresponding home
287 ferments, despite containing the same live active cultures (**Figure 3a, Figure S1a-S1b**).
288 PCoA analysis of tea samples, **Figure 3c**, revealed unambiguous differentiation of tea
289 leaves and brewed tea, as well as differences between tea types. Note, blank water samples
290 were most differentiated from the solid extract samples along PC1 (**Figure 3c**). Twelve
291 different teas were sampled at 0.5 min, 1 min, and 4 min to explore the brewing process with
292 respect to time and emulate tea that has been left steeping for longer periods of time (240
293 min). The tea samples, regardless of type, appeared most similar (in PCoA space and hence
294 chemistry) to the water blanks at the earliest time points and became more similar to solid
295 samples over time (along the PC1 axis which explained 25.9% of total variance), reflecting
296 the typical brewing process while measuring empirically the release of compounds from the
297 leaves (**Figure 3c, Figure 3d. Figure S1c**). The kinetics of tea extraction in the PCoA plot
298 shared similar trends for all teas. We observed minor chemical differences between 240 min
299 and 4 min for all tea types, which supports a steeping time rationale that appears to be
300 effective for extraction of phytochemicals from tea. Differences, based on tea type, were also
301 observed: white, green, matcha green, and black tea liquid samples were more similar to
302 each other than to oolong and pu'er, which were differentiated along PC3 (6.81% of total
303 variance) (**Figure 3c**); **Figure 3d and Figure S1c** illustrate clear differences in beta diversity
304 across different American and British teas and Chinese teas (oolong and pu'er), respectively.
305 It is noteworthy that, although PCoA enabled observation of overall trends, it did not display
306 changes in individual molecule concentrations. Individual chemical changes are visualized in
307 section 3.2 (Heatmaps) and discussed in section 3.3 (Molecular Networking and Annotations).

PCoA analysis of coffee revealed a clear trend among the sample types: brewed coffee (left hand side of PCoA) vs ground coffee (right hand side of PCoA) (**Figure S1d**). Besides samples clustering by sample type, different roast types could not be identified using PCoA but, as noted, changes in individual molecules are not captured using this approach. It is possible that molecular changes induced by roasting might be observed predominantly as volatile molecules, which were not assessed in this study. Indeed, changes in aroma, between the different roasting types, could be readily perceived.

Figure S1e-1f shows the effects spoilage had on ground beef and turkey, in the presence and absence of tetracycline, visualized along PC1, 20.88%. A key driver of chemical differences in PCoA was the type of meat (turkey or beef) (**Figure S1e; PC2 8.4%**). As tetracycline is used in the cattle industry, in some countries/ regions, we sought to identify if it also resulted in differences in chemical composition over time, possibly due to changes in microbial colonization and degradation during spoilage. To control for factors such as source and treatment of the meat, we added exogenous tetracycline as a treatment and the vehicle to controls. Samples with and without tetracycline changed similarly over time indicating that, based on the untargeted LC-MS/MS analysis, addition of tetracycline did not greatly impact chemical changes during aging processes (**Figure S1f)**.

PCoA analysis of the tomato samples revealed that both source (**Figure S1g**) and storage time (**Figure S1h**) affected the molecular composition of tomatoes. As expected, processed tomato samples (canned and sundried) occupied very different PCoA spaces than fresh tomatoes (**Figure S1g**). Chemical differences between the processed and fresh tomato could be attributed to the contribution of processing (e.g. heating or addition of sugar and oil) as well as packaging materials and not just the chemical compositions of tomatoes. It was

331 also notable that differences existed for fresh tomatoes, with those from the farmers' market
332 most closely resembling home-grown tomatoes, and all store-bought tomatoes resembling
333 one another, whether organic or not. It is likely that the close similarity of garden and farmers'
334 market roma tomatoes resulted from similar treatment, where the fruits are ripened on the
335 vine and collected and sold without processing, including storage in different types of packing
336 or washing (this is known for the garden tomatoes and presumed for the farmers' market).
337 Conversely, store-bought tomatoes are collected at an early stage, often not fully ripened for
338 ease of transportation, transported over long distances, packaged, and treated with
339 exogenous ethylene (depending on the supplier). This appears to have a more significant
340 effect on the chemical compositions of tomatoes than the "organic" designation. When
341 organic cherry tomatoes were left at room temperature to ripen, their molecular composition
342 changed over time (**Figure S1h**), although the tomatoes did not notably change in either
343 appearance or smell.

344 *3.2 Heatmaps for Identification of Chemical Changes by Food Group*

345 We created heatmaps to visualize molecular changes between samples for time
346 course experiments, specifically brewing tea, yogurt fermentation, tomato ripening, and
347 improper meat storage, and to gain insight into features that behave similarly over time or
348 originate from different sample types.

349 Complementary to a PCoA, heatmaps provide a visual overview of data to give more
350 detailed information about molecular changes driving differences within and between sample
351 types. **Figure 4** shows tea and milk-to-yogurt time courses, which had the largest changes in
352 abundance; heatmaps for other sample types are included in the Supplementary Information
353 (**Figure S2-S5**). Consistent with the PCoA analysis, we observed different metabolite profiles

between tea leaves and brewed tea (**Figure 4a**). Furthermore, we observed an increase in relative intensities of molecular features due to longer brewing times, independent of the tea type. We assessed the correlation of relative intensity per feature and tea type with extraction time, which resulted in a total of 2,045 significantly correlated features (*spearman correlation*, *p-value* < 0.05). For example, we observed that the relative intensities of procyanidin B and theaflavin increased over time (*Kruskal–Wallis, N=6*, *p-value ranging from 0.01 to 0.02, between brewing times 0.5 and 240*) (**Figure 4b and Figure S6a**). We also assessed the correlation of relative intensity per feature and home ferment with different yogurt inoculums over time. For the Kroger yogurt, this resulted in a total of 1,587 significantly correlated features (*spearman correlation*, *p-value* < 0.05) (**Figure 4c**). **Figure 4d and Figure S6b** highlight selected molecular features for which we obtained putative structure annotations through GNPS library matching. For example, we observed that the relative intensity of 4-O-beta-galactopyranosyl-D-mannopyranose decreased over time for each yogurt type individually as well as overall (*Kruskal–Wallis, N=9*, *p-value=0.0023*, *between 0 and 58 hrs*) **(Figure 4d)**.

Molecular changes during meat (beef and turkey) storage over five days were also visualized using a heat map (**Figure S3a**). When comparing antibiotic- vs non-antibiotic-treated meat (beef and turkey), overall molecular differences, as seen in PCoA space, did not vary. However, there were some specific low intensity molecules that changed, although differences were minimal due to the addition of tetracycline, which was consistent with observations from the PCoA. We observed differences between organic and non-organic beef. For example, in non-organic beef, oleoyl-taurine increased during the 5 days but did not appear in organic samples, while concentrations acetyl-carnitine decreased in non-organic

beef but were consistent across all time points for organic beef. In turkey, the rates of appearance of oleoyl-taurine and disappearance of acetyl-carnitine were only slightly different (**Figure S3b**). The spectral match to the fungal molecule, termitomycamide E (Choi et al., 2010), with parent mass difference 0.000 Da and very strong cosine match of 0.84, increased over time. The presence of three analogues, with mass differences pointing to different acyl chain lengths, and suppression with the addition of tetracycline, would be consistent with increased microbial (fungal) loads (**Figure S5**).

Molecular differences between tomato samples were most striking when comparing sun dried, canned, and fresh tomatoes. In the heatmap, no clear-cut large-scale patterns were observed when visualizing molecular changes during ripening of fresh tomatoes (**Figure S2**). During the ripening process, some individual molecular features were found to decrease in relative abundance. For example, 5'-methylthioadenosine, a key ripening hormone for plants and precursor of plant-produced ethylene (North et al., 2017), was found to have decreased significantly in relative abundance over the 5-day time course. Also, plant flavonoids (including a level-3 annotation for naringenin) and tomatidine, a tomato-specific alkaloid (Brink and Folkers, 1951; Friedman, 2013), were found to have decreased significantly in relative intensity over time. This is informative as many of the health properties associated with consumption of polyphenol-containing foods are attributed to molecules like naringenin and our results, therefore, indicated tentatively that the nutritional value of tomatoes might change over the time period tomatoes are stored in the home environment.

3.3 Molecular networking and identification of known and related compounds

Mass spectral molecular networking provided additional information about molecular relationships that complemented global differences captured by PCoA and further explored

400 the molecules and molecular changes within each food type. In milk and yogurt, matches with
401 six carbon sugars, disaccharides and oligosaccharides, vitamins, and acylated carnitines
402 were observed (**Figure 5b**). In addition, large lipid molecular families, such as sphingolipids,
403 and glycerol conjugated with fatty acids, such as monoolein and linoleoylglycerol, were
404 identified. Delvocid, also known as the clinical antifungal 'natamycin', which is an additive
405 used to preserve dairy products, was detected (Branen et al., 2001) and did not change in
406 relative abundance over time. These annotations are all consistent with the animal origins of
407 the samples (milk, yogurt). However, we also obtained unexpected annotations. A molecular
408 family of bile acids, containing annotated glycocholic and cholic acids, was identified. This
409 molecular family was not expected to be present in these samples, as they are primarily
410 associated with the gut. Putative assignments were inspected and confirmed using manual
411 inspection of raw data, accurate mass, fragmentation pattern, and retention time analysis,
412 which further supported the presence of this molecular family.
413 A large range of phytochemicals were annotated in tea samples (**Figure 6c and**
414 **Supplementary Figure 12**), including large molecular families associated with flavonoids,
415 with spectral matches to puerins, catechins, and apigenin (assignments are putative as
416 isomers are difficult to differentiate in accordance with level 3 metabolite identification
417 (Cuyckens & Claeys, 2004, Sumner et al., 2007, Borges et al., 2018). MS work on tea has
418 been done primarily in negative ionization mode. Here, using the positive ionization mode, we
419 corroborated earlier work finding molecular families containing flavonoid aglycones with
420 MS/MS matches to quercetin, kaempferol, myricetin, and (epi)catechin – glycosides of which
421 are abundant in tea (van der Hooft et al., 2012a) – and a large molecular family consisting of
422 glycoside derivatives that had spectral matches with quercetin and kaempferol that were

423 bundled together with chlorogenic acids. Note, the majority of nodes for this family were
424 annotated with GNPS community contributed library hits, indicating a greater library coverage
425 for some compound classes, likely due to community contributed spectra. As expected,
426 caffeine was also annotated in tea samples. Theaflavin, a polyphenol formed during fungal
427 oxidation and its analogues, often associated with black tea (Zhang et al., 2018), were
428 detected in white, green, black and oolong tea samples. Theaflavin increased in relative
429 concentration over time, as shown in **Figure 4 and Figure S13**. These annotations were
430 consistent with known processes that use polyphenol building blocks to create larger
431 scaffolds like theaflavin and give black tea its typical color. Furthermore, fuzhuanins,
432 polyphenol-derived molecules (Luo et al., 2013), which are beta-ring fission lactones of
433 flavan-3-ols like epicatechin, were found at high abundance in tea samples.
434 In coffee (**Figure 5c**), we also observed caffeine as well as methyl-caffeine (1,3,7,8-
435 Tetramethylxanthine) and another related compound with a delta mass of *m/z* 14.01 (CH_2),
436 corresponding to theobromine. Furthermore, we detected several flavonoids and a large
437 number of hydroxycinnamic acids and chlorogenic acids, which are commonly observed in
438 plants (Islam et al., 2018; Clifford et al., 2017; Pastoriza et al., 2017; Tajik et al., 2017;
439 Naveed et al., 2018). In addition, library matching revealed the presence of mascarosides,
440 molecules commonly observed upon roasting of coffee (Shu et al., 2014). The mascarosides
441 were identified in the molecular network by *m/z* 162.053, 15.996 and 18.011 gains and
442 losses, corresponding to mass shifts associated with six carbon sugars, oxygen, and water,
443 respectively – all pointing consistently to the presence of glycosylated mascarosides.
444 In the meat samples (**Figure 6a)**, as expected, we observed MS/MS matches to
445 tetracycline displayed as a single node (no related spectra were detected), which were more

abundant in turkey samples. Although tetracycline is used commonly as a growth promoter, here it was added to see the effects of this antibiotic on a 5-day food spoilage test (Granados-Chinchilla & Rodríguez, 2017). We also observed spectral matches with carnosine as well as a large cluster of acyl carnitines with five spectral matches to different acylations. The acyl carnitines were observed predominantly in beef. However, we also found a molecular family of N-acyltaurines (NATs), a recently discovered class of lipids (Turman et al., 2008). **Figure S3** shows how NAT concentrations increased, after two days of storage at room temperature, whereas levels of acylcarnitines (markers for beta-oxidation) dropped, suggesting these chemical changes were associated with decomposition over time. Ceramides, component lipids in eukaryotic cell membranes, were detected in both beef and turkey, but they only fell below the level of detection after 5 days in beef. Their presence suggested disintegration of cells within the tissue and the lability of ceramides would explain their disappearance over time. Oxidation in the presence of haem-iron may have contributed to increased degradation in red meat compared with turkey. In contrast, a molecular family containing carnosol, a metabolite from rosemary (Loussouarn et al., 2017), was observed in turkey, but not in beef. Only the packaging of turkey grown without antibiotics and growth hormones stated that rosemary was used (see study Metadata), yet it was observed in both conventionally grown as well as antibiotic-free meat. Both dipeptides and *N*-methyl histidine were detected during the 5-day aging process of the meat. Thus, the molecular families described here are consistent with these sample types. Additionally, the changes observed in chemical composition, emerging after 2 days of storage for a relatively small number of molecular features, suggested these might be used as signature compounds for decomposition in meat samples (Supplementary Figure 2).

469 Many known chemicals in tomatoes were detected, including chlorogenic acid
470 derivatives and flavonoids **(Figure 5a)**, both compound groups are found commonly in tomato
471 cultivars (van der Hooft et al., 2012b; Floros et al., 2017). A molecular family of tomatidine-
472 related molecules was observed in all tomato samples. Tomatidine, a tomato-specific alkaloid,
473 as the name suggests, and structural component of related glycoalkaloids, is abundant in
474 tomato plant leaves and stems but is less concentrated in the fruits. Similarly, phenylethyl
475 pyranosides were observed in all tomatoes. Only in sundried tomatoes did we observe a
476 spectral match with glucose, perhaps added as a sweetener. In both sundried and fresh
477 tomatoes, we detected azoxystrobin, a fungicide used in agriculture. Many molecules,
478 including added oils, sugars, and preservatives, might explain the differences observed
479 between processed and raw tomatoes in the PCoA (Figure S1g, Figure 5a). 5'-
480 methylthioadenosine, one source of ethylene, a ripening hormone in plants, was detected in
481 all tomatoes, except sundried tomatoes (North et al., 2017). The relative concentrations of 5'-
482 methylthioadenosine were observed to decrease over time/ripening (**Figure S2).**
483
484 *3.4 Molecular transformations*
485 Heatmaps as well as molecular networks, while very different but complementary
486 visualization techniques, confirmed and complemented one another and provide additional
487 perspectives to augment current food analysis. For example, theaflavin increased significantly
488 in relative abundance over time, which was visualized in the heatmap (**Figure 4a,b**) as well
489 as the molecular network displaying brewing time (**Figure S13**). In **Figure 6c**, theaflavin is
490 associated with two unannotated compounds, which allowed the presence of two analogues
491 with mass shifts of 16 and 30 Da to be confirmed and are consistent with a hydroxylated as

492 well as a double de-hydroxylated analogues, and other further reduced analogues. These
493 kinds of relationships facilitate interpretation and understanding of chemical processes
494 without requiring the identity of molecules detected to be known. The molecular composition
495 of tea samples changed over time, with changes observed in the abundance consistent with
496 continued extraction of molecules as opposed to chemical modifications. While a range of
497 compounds increased in many of the varieties, there were molecular features specific to tea
498 type, such as increased relative abundance of coniferyl aldehyde in only oolong tea (**Figure**
499 **S6a** panel 2 and **Figure 6c** cluster 8). The molecular composition and extraction kinetics of
500 oolong tea might differ from other varieties as a result of extensive drying, physical changes in
501 the leaves (*e.g.* twisting/curling), and oxidation during production.
502 The changes observed in tea samples were in contrast to the yogurt samples, where
503 chemical alterations over time varied significantly, likely due to microbial metabolism. We
504 detected significant changes in PCoA, molecular networks, and heatmaps. In the PCoA, the
505 home ferment inoculated with Kroger yogurt resembled the original starting culture, at a
506 molecular level, and differed from other home ferments, possibly because it contained
507 different yogurt cultures. Interestingly, when we focused our analysis on annotated
508 compounds only, significant changes over time were not observed in the heatmap **(Figure 4**
509 and **Figure S4)**, indicating that many of the molecular transformations during fermentation
510 have not been characterized yet or the reference spectra are not available in MS library
511 databases. Consistent with the lack of reference spectra in public databases, the yogurt and
512 milk samples also had the lowest annotation rate at 3.5%. Among the annotated features, we
513 found a broad range of compounds **(Figure S8)**, including food additives and sugars, which
514 were also detected in other milk types within publicly available datasets on GNPS, such as

breast milk. The unexpected occurrence of bile acids in the milk and yogurt samples might originate from secretion from the cows' mammary glands into the milk, as bile acids have been detected previously in human breast milk (Forsyth et al., 1983).

4. Conclusion

Our study presents the first large-scale food composition analysis using mass spectral molecular networking. The untargeted MS approach coupled with molecular networking allowed us to assess large-scale differences between sample types, find molecule-molecule links within and between sample types, and identify different compound classes found within a sample type - all useful in biochemical interpretations and understanding. We determined that foods undergo molecular changes caused by a variety of biological and chemical processes over different time periods, as exemplified by meat, tea and yogurt. Brewing time for tea altered its composition, increasing the diversity of molecules, whereas fermentation of yogurt from milk, spoilage of meat, and ripening of tomatoes were all dominated by biological transformations, altering the molecular composition over a longer period. Mass spectral molecular networking and spectral library search successfully identified key molecular features, which differed based on processing type, such as fermentation time in the yogurt samples and brewing time for tea. Our study provides a reference dataset freely accessible for feature mining in future food-related or other studies. One advantage of the GNPS molecular networking workflow is the search parameter 'Find Related Datasets'. As shown in this study, even the most traditional food types contain large numbers of unannotated molecules and, therefore, we expect that increasing depositions of MS datasets in the public domain would allow comparisons with other complex mixtures and narrow down the origins of

molecular features. This is the first large-scale food chemistry study, free and publicly accessible through the KnowledgeBase GNPS (Wang et al., 2016). Anyone who wishes to continue exploring these data can subscribe to the project, as it will be subject to living data analysis. Living data is a strategy introduced to metabolomics in Wang et al. (2016), where data are continuously re-analyzed, and updates are provided automatically to all subscribers. This allows future studies to exploit the mass spectral molecular networking data with the annotated food molecules, with the ability to propagate annotations across a new network to better understand the chemical space that foods occupy and how food handling and processing affect it. Given that 88-97% of all the MS/MS spectra are currently unannotated, as a community, we will need to increase our knowledge about the molecular compositions and molecular changes in our food. We have shown that, with our contribution to the food chemistry field, GNPS molecular networking promises to become a key repository and knowledgebase for untargeted MS-based food composition studies, and demonstrated the utility of combining molecular networking approaches with statistical measures to discern meaningful chemical transformations.

Acknowledgments

The result of this work was a part of a hands-on mass spectrometry course at UCSD called "System Wide Mass Spectrometry". The authors were all participants or mentors of this course. We further acknowledge NIH number P41 GM103484, and NIH Grant GMS10RR029121 and Bruker for the shared instrumentation infrastructure that enabled this work and the UCSD Center for Microbiome Innovation. AMCR and PCD were supported by NSF grant IOS-1656481. FV and PCD were supported by the Office of Naval Research

Multidisciplinary University Research Initiative (MURI) Award, Award number N000014-15-1-2809.

Appendix

Supplementary data

Supplementary data associated with this article can be found with the online version.

Data and code availability

Data can be accessed via http://gnps.ucsd.edu via accession numbers for Milk/yogurt: MSV000082387; Tea: MSV000082388; Coffee: MSV000082386; Meat: MSV000082423; Tomato: MSV000082391. Metadata are uploaded with each dataset. All jupyter notebooks and scripts used for data pre-processing and analysis are publically available at: https://github.com/DorresteinLaboratory/supplementary-MolecularChangesInFood

Links to networking jobs for GNPS networking jobs: Parameters are precursor and fragment ion tolerance set to 0.1 Da, min matched fragment ions: 4, cosine 0.7; library search min match peaks: 4; run with metadata, attributes assigned; NIST17 included.

milk/yogurt samples only

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=1f11fb76e15240a893e46e02d9c58cd2

tea samples only

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=f7569832b0d241f79812446d81cd5ca5

coffee samples only

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=2c83502fefc0469ba79ba70a9461b9b5

meat samples only

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=1520f7112d9f445384eb743ac4358c21

tomato samples only

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=cf6e0347de8b48aeb8c700e45b4d3159

All sample types:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=b881151839574f639ceaf06f9b11e464

Links for jobs performed to obtain the feature MS1 and MS/MS linked table for PCoA and heatmaps and statistical analysis: Parameters are precursor and fragment ion tolerance set to 0.1 Da, 4 min match peaks, cosine 0.7; inputs are the .mgf; .csv from mzMINE and the metadata file.

Global:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=25577f11a35c48cdb30dfc005cbd6638

milk/yogurt:

http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=f9ede96d9e694f9e8b6186991e289c17

tea:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=80fc6384e52b4fe18e13ebbb3a86b4d8

coffee:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=750cc6fe82dc4732b84b355904cf91d3

meat:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=85c7380ca7ee44b3a3ed448c9b4d09fa

tomato:

http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=7a4031dd3ee146699ceecef919d7f668

Link to Clusterapp:

http://dorresteinappshub.ucsd.edu:3838/clusterMetaboApp0.9.1/

Conflict of interest statement

Dorrestein is on the advisory board of Sirenas. NB was a co-founder, had an equity interest and received income from Digital Proteomics, LLC through 2017. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. Digital Proteomics was not involved in the research presented here.

References

Aksenov, A. A., da Silva, R., Knight, R., Lopes, N. P., & Dorrestein, P. C. (2017). Global chemical analysis of biology by mass spectrometry. Nature Reviews Chemistry, 1, 0054.

Berger, J., Roos, N., Greffeuille, V., Dijkhuizen, M., & Wieringa, F. (2019). Driving Policy Change to Improve Micronutrient Status in Women of Reproductive Age and Children in Southeast Asia: The SMILING Project. *Maternal and Child Health Journal*, *23*(1), 79–85.

Blaženović, I., Kind, T., Ji, J., & Fiehn, O. (2018). Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites*, 8(2), 31.

Borges, R. M., Taujale, R., de Souza, J. S., de Andrade Bezerra, T., Silva, E. L. E., Herzog, R., Ponce, F. V., Wolfender, J. L., & Edison, A. S. (2018). Dereplication of plant phenolics using a mass-spectrometry database independent method. *Phytochemical Analysis*, (in press) doi:10.1002/pca.2773.

Branen, A. L., Davidson, P. M., Salminen, S., & Thorngate, J. (2001). Food Additives. New York: CRC Press. pp. 599–600. ISBN 9780824741709.

Brink, N. G., & Folkers, K. (1951). Isolation of Tomatidine from the Roots of the Rutgers Tomato Plant. *Journal of the American Chemical Society*, *73*(8), 4018.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N, Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., & Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5), 335–336.

Choi, J. H., Maeda, K., Nagai, K., Harada, E., Kawade, M., Hirai, H., & Kawagishi, H. (2010). Termitomycamides A to E, fatty acid amides isolated from the mushroom *Termitomyces titanicus*, suppress endoplasmic reticulum stress. *Organic Letters*, 12(21), 5012–5015.

640 Clifford, M. N., Jaganath, I. B., Ludwig, I. A., & Crozier, A. (2017). Chlorogenic acids and the
641 acyl-quinic acids: discovery, biosynthesis, bioavailability and bioactivity. *Natural Product*
642 *Reports,* 34(12), 1391–1421.
643 Cuyckens, F. & Claeys, M. (2004). Mass spectrometry in the structural analysis of flavonoids.
644 *Journal of Mass Spectrometry*, 39(1), 1–15.
645 Floros, D. J., Petras, D., Kapono, C. A., Melnik, A. V., Ling, T.-J., Knight, R., & Dorrestein,
646 P. C. (2017). Mass spectrometry based molecular 3D-cartography of plant metabolites.
647 *Frontiers in Plant Science*, 8, 429.
648 Forsyth, J. S., Ross, P. E. & Bouchier, I. A. D. (1983). Bile salts in breast milk. Eur J Pediatr,
649 140: 126-127.
650 Friedman, M. (2013). Anticarcinogenic, Cardioprotective, and Other Health Benefits of
651 Tomato Compounds Lycopene, α-Tomatine, and Tomatidine in Pure Form and in Fresh
652 and Processed Tomatoes. *Journal of Agricultural and Food Chemistry*, *61*(40), 9534–
653 9550.
654 Ge, Y. W., Zhu, S., Yoshimatsu, K., & Komatsu, K. (2017). MS/MS similarity networking
655 accelerated target profiling of triterpene saponins in *Eleutherococcus senticosus* leaves.
656 *Food Chemistry*, 227, 444–452.
657 Giorio, C., Safer, A., Sánchez-Bayo, F., Tapparo, A., Lentola, A., Girolami, V., van Lexmond,
658 M. B., & Bonmatin, J.M. (2017). An update of the Worldwide Integrated Assessment (WIA)
659 on systemic insecticides. Part 1: new molecules, metabolism, fate, and transport.
660 *Environmental Science and Pollution Research*. (in press) doi: 10.1007/s11356-017-
661 0394-3.
662 Granados-Chinchilla, F. & Rodríguez, C. (2017). Tetracyclines in food and feeding stuffs:
663 From regulation to analytical methods, bacterial resistance, and environmental and health
664 implications. *Journal of Analytical Methods in Chemistry*, 2017, 1315497.
665 Islam, M. T., Tabrez, S., Jabir, N. R., Ali, M., Kamal, M. A., da Silva Araujo, L., De Oliveira
666 Santos, J. V., Da Mata, A. M. O. F., De Aguiar, R. P. S., & de Carvalho Melo Cavalcante,
667 A. A. (2018). An insight on the therapeutic potential of major coffee components. *Current*
668 *Drug Metabolism*, (in press) doi: 10.2174/1389200219666180302154551.
669 Luo, Z. M., Du, H. X., Li, L. X., An, M. Q., Zhang, Z. Z., Wan, X. C., Bao, G. H., Zhang, L., &
670 Ling, T. J. (2013). Fuzhuanins A and B: the B-ring fission lactones of flavan-3-ols from
671 Fuzhuan brick-tea. *Journal of Agricultural and Food Chemistry*, 61(28), 6982–6990.
672 Lewis, J. E., Arheart, K. L., LeBlanc, W. G., Fleming, L. E., Lee, D. J., Davila, E. P., … Clark,
673 J. D. (2009). Food label use and awareness of nutritional information and
674 recommendations among persons with chronic disease. *The American journal of clinical*
675 *nutrition*, *90*(5), 1351–1357.
676 Marshall, J. W., Schmitt-Kopplin, P., Schuetz, N., Moritz, F., Roullier-Gall, C., Uhl, J., Colyer,
677 A., Jones, L. L., Rychlik, M., & Taylor, A. J. (2018). Monitoring chemical changes during
678 food sterilisation using ultrahigh resolution mass spectrometry. *Food Chemistry*, 242,
679 316–322.

680 Naveed, M., Hejazi, V., Abbas, M., Kamboh, A. A., Khan, G. J., Shumzaid, M., Ahmad, F.,
681 Babazadeh, D., FangFang, X., Modarresi-Ghazani, F., WenHua, L., & XiaoHui, Z. (2018).
682 Chlorogenic acid (CGA): A pharmacological review and call for further research.
683 *Biomedicine & Pharmacotherapy*, 97:67–74.
684 Nguyen, D. D., Wu, C. H., Moree, W. J., Lamsa, A., Medema, M. H., Zhao, X., Gavilan, R.
685 G., Aparicio, M., Atencio, L., Jackson, C., Ballesteros, J., Sanchez, J., Watrous, J. D.,
686 Phelan, V. V., van de Wiel, C., Kersten, R. D., Mehnaz, S., De Mot, R., Shank, E. A.,
687 Charusanti, P., Nagarajan, H., Duggan, B. M., Moore, B. S., Bandeira, N., Palsson, B. Ø.,
688 Pogliano, K., Gutiérrez, M., & Dorrestein, P. C. (2013). MS/MS networking guided analysis
689 of molecule and gene cluster families. *Proceedings of the National Academy of Sciences*
690 *of the United States of America*, 110(28), E2611–E2620.
691 North, J. A., Miller, A. R., Wildenthal, J. A., Young, S. J., & Tabita, F. R. (2017). Microbial
692 pathway for anaerobic 5′-methylthioadenosine metabolism coupled to ethylene formation.
693 *Proceedings of the National Academy of Sciences of the United States of America*,
694 114(48), E10455–E10464.
695 Pastoriza, S., Mesías, M., Cabrera, C., & Rufián-Henares, J. A. (2017). Healthy properties
696 of green and white teas: an update. *Food & Functions*, 8(8), 2650–2662.
697 Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C., Dragsted, L. O., Draper, J.,
698 Rappaport, S. M., van der Hooft, J. J. J., & Wishart, D. S. (2014). The food metabolome:
699 a window over dietary exposure. *American Journal of Clinical Nutrition*, 99(6), 1286–1308.
700 Scheubert, K., Hufsky, F., Petras, D., Wang, M., Nothias, L. -F., Dührkop, K., Bandeira, N.,
701 Dorrestein, P. C., & Böcker, S. (2017). Significance estimation for large scale
702 metabolomics annotations by spectral matching. *Nature Communication*, 8(1), 1494.
703 Shu, Y., Liu, J. Q., Peng, X. R., Wan, L. S., Zhou, L., Zhang, T., & Qiu, M. H. (2014).
704 Characterization of diterpenoid glucosides in roasted puer coffee beans. *Journal of*
705 *Agricultural and Food Chemistry*, 62(12), 2631–2637.
706 Schulze, M. B., Manson, J. E., Ludwig, D. S., Colditz, G. A., Stampfer, M. J., Willett, W. C.,
707 … Gimble, J. M. (2016). 2015 – 2020 Dietary Guidelines for Americans. *Am J Clin Nutr*.
708 Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W.-
709 M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi,
710 R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., Reily, M. D., Thaden,
711 J. J., & Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis
712 Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).
713 *Metabolomics*, 3(3), 211–221.
714 Tajik, N., Tajik, M., Mack, I., & Enck, P. (2017). The potential effects of chlorogenic acid, the
715 main phenolic components in coffee, on health: a comprehensive review of the literature.
716 *European Journal of Nutrition*, 56(7), 2215–2244.
717 Thirumdas, R., Brnčić, M., Brnčić, S. R., Barba, F. J., Gálvez, F., Zamuz, S., … Lorenzo, J.
718 M. (2018). Evaluating the impact of vegetal and microalgae protein sources on proximate

composition, amino acid profile, and physicochemical properties of fermented Spanish "chorizo" sausages. *Journal of Food Processing and Preservation*, 42 (11), e13817

Turman, M. V., Kingsley, P. J., Rouzer, C. A., Cravatt, B. F., & Marnett, L. J. (2008). Oxidative metabolism of a fatty acid amide hydrolase-regulated lipid, arachidonoyltaurine. *Biochemistry*, 47(12), 3917–3925.

van der Hooft, J. J. J., Akermi, M., Ünlü, F. Y., Mihaleva, V., Roldan, V. G., Bino, R. J., de Vos, R. C., & Vervoort, J. (2012a). Structural annotation and elucidation of conjugated phenolic compounds in black, green, and white tea extracts. *Journal of Agricultural and Food Chemistry*, 60(36), 8841–8850.

van der Hooft, J. J. J., Vervoort, J., Bino, R. J., & de Vos, R. C. H. (2012b). Spectral trees as a robust annotation tool in LC–MS based metabolomics. *Metabolomics*, 8(4), 691–703.

Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., & Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. *Gigascience*, 2(1), 16.

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W. T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R. D., Pace, L. A., Quinn, R. A., Duncan, K. R., Hsu, C. C., Floros, D. J., Gavilan, R. G., Kleigrewe, K., Northen, T., Dutton, R. J., Parrot, D., Carlson, E. E., Aigle, B., Michelsen, C. F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B. T., Gerwick, L., Liaw, C. C., Yang, Y. -L., Humpf, H. U., Maansson, M., Keyzers, R. A., Sims, A. C., Johnson, A. R., Sidebottom, A. M., Sedio, B.E., Klitgaard, A., Larson, C. B., Boya P, C. A., Torres-Mendoza, D., Gonzalez, D. J., Silva, D. B., Marques, L. M., Demarque, D. P., Pociute, E., O'Neill, E. C., Briand, E., Helfrich, E. J. N., Granatosky, E. A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J. J., Zeng, Y., Vorholt, J. A., Kurita, K. L., Charusanti, P., McPhail, K. L., Nielsen, K. F., Vuong, L., Elfeki, M., Traxler, M. F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R. R., Mascuch, S. J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P. G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A. M. C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B. M., Almaliti, J., Allard, P. M., Phapale, P., Nothias, L. -F., Alexandrov, T., Litaudon, M., Wolfender, J. L., Kyle, J. E., Metz, T. O., Peryea, T., Nguyen, D. T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K. M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P. R., Palsson, B. O., Pogliano, K., Linington, R. G., Gutiérrez, M., Lopes, N. P., Gerwick, W. H., Moore, B. S., Dorrestein, P. C., & Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, 34(8), 828–837.

Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J. M., Moore, B. S., Laskin, J., Bandeira, N., & Dorrestein, P. C. (2012). Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences of the United States of America*, 109(26), E1743–E1752.

759 Yang, J. -Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., Glukhov, E.,
760 Wodtke, A., de Felicio, R., Fenner, A., Wong, W. R., Linington, R. G., Zhang, L., Debonsi,
761 H. M., Gerwick, W. H., & Dorrestein, P. C. (2013). Molecular networking as a dereplication
762 strategy. *Journal of Natural Products*, 76(9), 1686–1699.
763 Yoshimura, Y., Goto-Inoue, N., Moriyama, T., & Zaima, N. (2016). Significant advancement
764 of mass spectrometry imaging for food chemistry. *Food Chemistry*, 210, 200–211.
765 Zhang, S., Yang, C., Idehen, E., Shi, L., Lv, L., & Sang, S. (2018). Novel theaflavin-type
766 chlorogenic acid derivatives identified in black tea. *Journal of Agricultural and Food*
767 *Chemistry,* 66(13), 3402-–3407.
768

769

770 **Figure captions**

771

772 **Figure 1.** Representative images of foods sampled and timeline of sampling. From top to
773 bottom: Yogurt preparation from milk, brewed and loose leaf tea, different coffees,
774 representing diverse roasts, brands and origins as ground coffee and brewed, ground beef
775 and ground turkey left out to spoil, image from 3 days of storage depicted on the far right, and
776 tomato types and ripening timeline. RT denotes room temperature and // denotes a time
777 break. [color reproduction]

778

779 **Figure 2.** Global PCoA analysis to understand the molecular relationships among all samples
780 analyzed. PC1 (10.43%); PC2 (5.96%); PC3 (5.86%). As a 2D image, the PCoA plot does not
781 reveal the relationships clearly, a movie rotating this image is provided as supporting
782 information [uploaded on massive.ucsd.edu MSV000083014]. MS1 features are TIC
783 normalized per sample and the PCoA analysis was performed using Qiime1 and the
784 Canberra distance metric. [color reproduction]

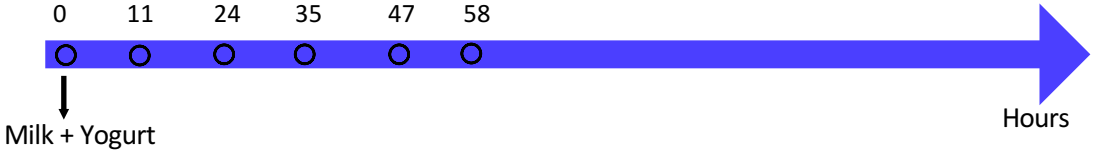
785

Figure 3. PCoA plots for the individual food types, color coded by metadata categories to visualize key drivers in molecular patterns. The three store bought yogurts containing live active cultures, the milk and the home ferments using the different yogurts as starter culture show distinct groupings. The spheres are colored based on milk and yogurt type (a) or fermentation time from 0 to 58 hrs (b). Tea samples differentiated based on tea type and brewed tea vs. tea leaves (c). The time course of tea extraction is displayed for American and British teas: black, green, matcha green, and white teas (d). [color reproduction]

Figure 4. Metabolites changing over tea extraction time and during the fermentation process from milk to yogurt. a) Heatmap showing tea metabolites changing over extraction time across different tea types. b) Specific metabolites increase significantly in their relative intensity during tea extraction time. c) Heatmap showing metabolites changing during the fermentation process from milk to yogurt across different yogurt brands used as inocula, as well as the milk as control. d) Metabolites increasing or decreasing significantly during the fermentation process across different home ferments. Metabolite annotation was performed through mass spectral molecular networking and spectral matching to reference spectra. [color reproduction]

Figure 5. Molecular network clusters of the a) tomato color coded by processing method, b) milk to yogurt, c) coffee data. The clusters are enlarged regions of specific molecular families observed within the full molecular network. The color coding for different samples groups are explained in the figure legend. Node sizes indicated relative precursor abundance and selected library identifications are annotated in the figure and shown through squared node

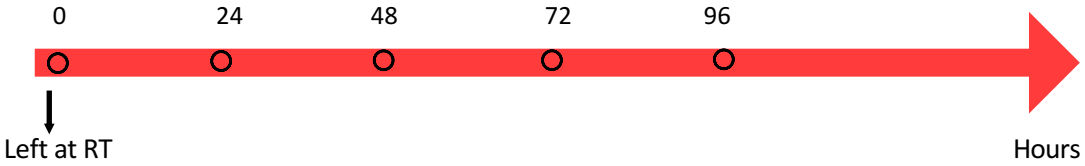
809 shape. The full size images of the entire network where one can zoom in to the molecular
810 networks can be found as supporting information (**Figure S7-S9**) and the GNPS links to the
811 analysis jobs are provided in the data availability section. All annotations shown are level 2 or
812 3 according to the 2007 metabolomics standards guidelines (Sumner et al., 2007). [color
813 reproduction]
814
815 **Figure 6.** Molecular networks of the data. a) reflect the meat samples color coded by turkey
816 or beef. b) same network as a) but color coded by aging time. c) molecular networks color
817 coded by tea. The clusters are enlarged regions of specific molecular families observed within
818 the full molecular network. The full size images of the entire molecular networks where one
819 can zoom in molecular networks can be found as supporting information (**Figures S10-S12**)
820 and the GNPS links to the analysis jobs are provided in the data availability section. All
821 annotations shown are level 2 or 3 according to the 2007 metabolomics standards guidelines
822 (Sumner et al., 2007). [color reproduction]
823

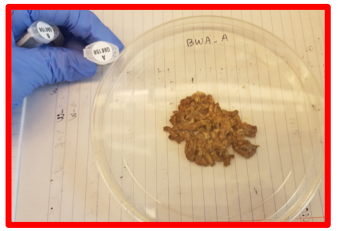


1 whole milk (Horizon Organic Vitamin D Milk)
3 brands of yogurt: Oikos plain Greek nonfat , Voskos plain Greek , Kroger plain nonfat Greek

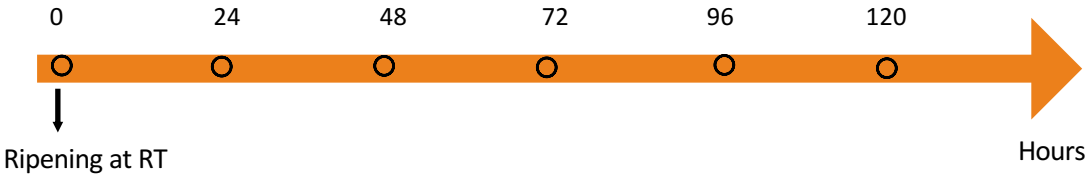
6 tea types:
White, Green, Pu'er
Matcha Green ,
Oolong, Black

38 varieties of coffee
different roast, brand and origin

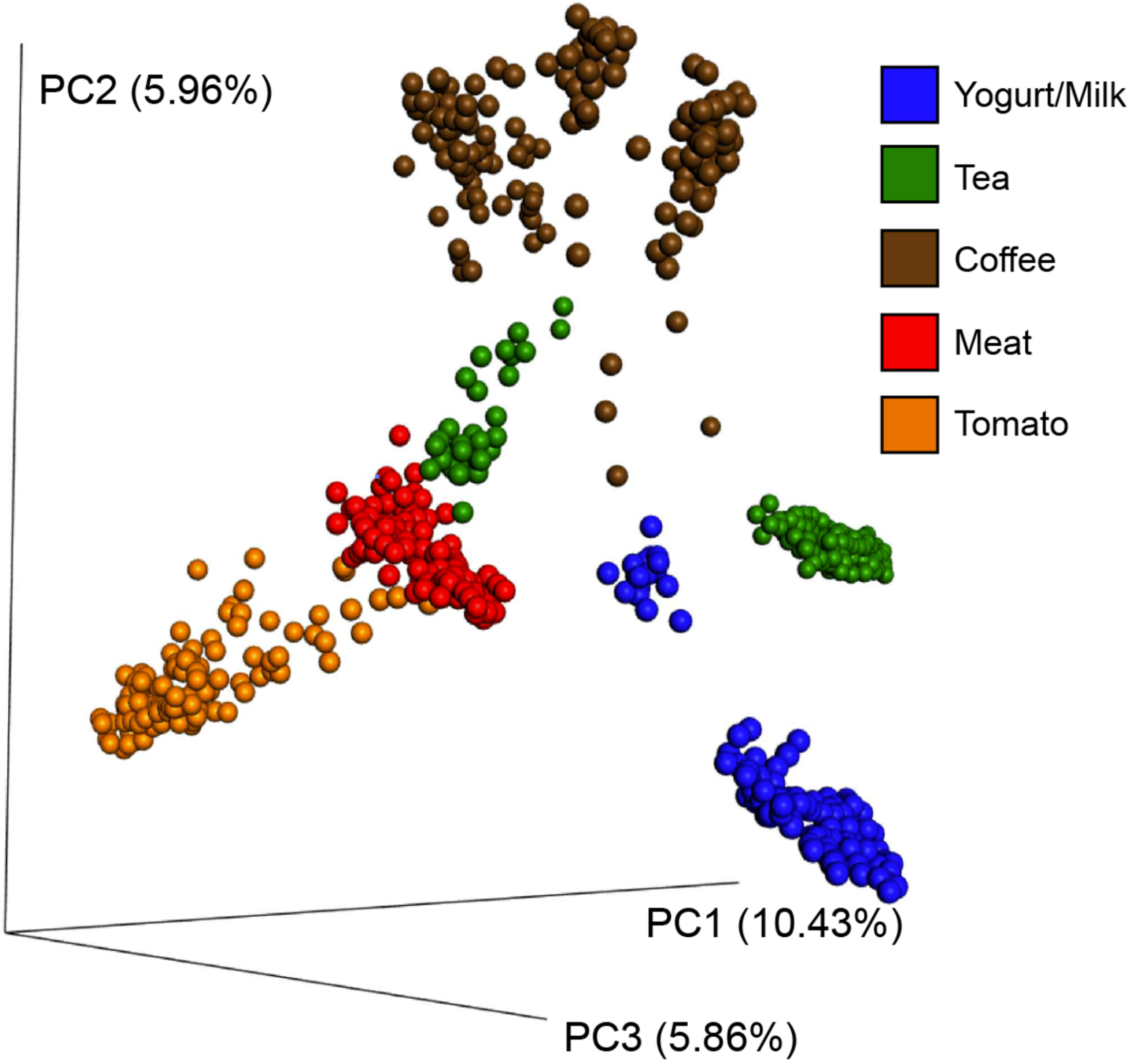




4 types of meat :
Ground Beef
Ground Beef (certified organic)
Ground Turkey
Ground Turkey (certified organic)



Organic cherry tomatoes
Non organic cherry tomatoes
Sun-dried tomatoes
Canned tomatoes


PC2 (5.96%)
Yogurt/Milk
Tea
Coffee
Meat
Tomato
PC1 (10.43%)
PC3 (5.86%)

a)

Voskos yogurt

Kroger yogurt

Oikos yogurt

Milk

PC1 (22.72%)

PC2 (15.39%)

PC3 (8.32%)

Horizon milk

Kroger yogurt

Oikos yogurt

Voskos yogurt

b)

Milk + Kroger yogurt

PC1 (22.72%)

PC2 (15.39%)

PC3 (8.32%)

Fermentation time [h]

0

11

24

35

47

58

c)

Brewed tea

Tea leaves

PC1 (25.99%)

PC2 (12.31%)

PC3 (6.81%)

black

green

matcha green

white

oolong

puer

water

d)

American and British

Brewed tea

Tea leaves

PC1 (25.99%)

PC2 (12.31%)

PC3 (6.81%)

Brewing time [min]

0.5

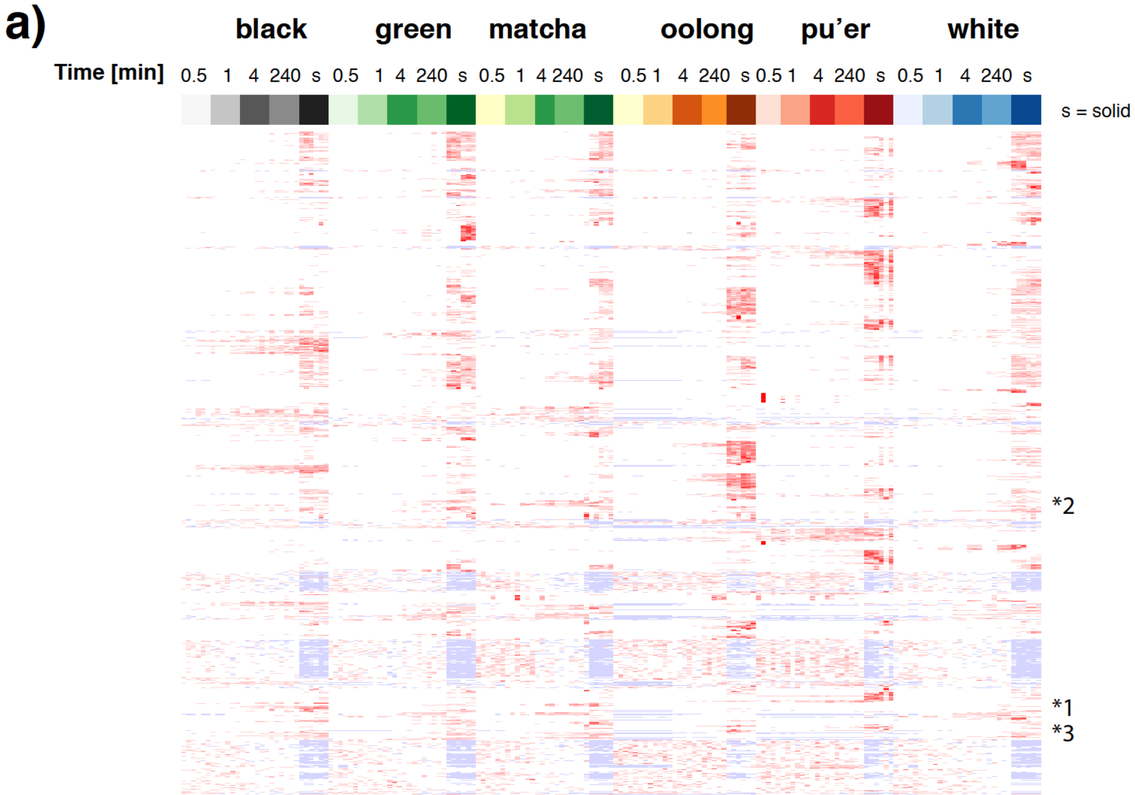
1

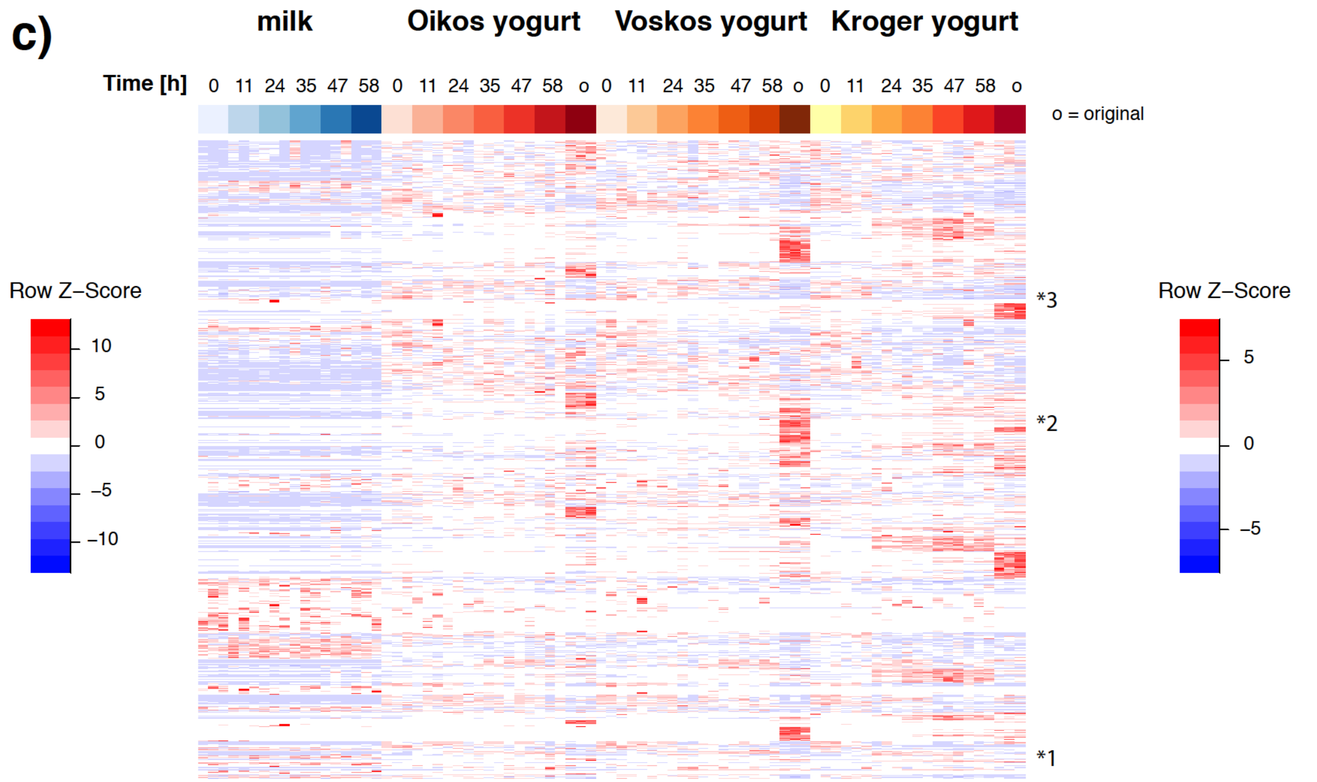
4

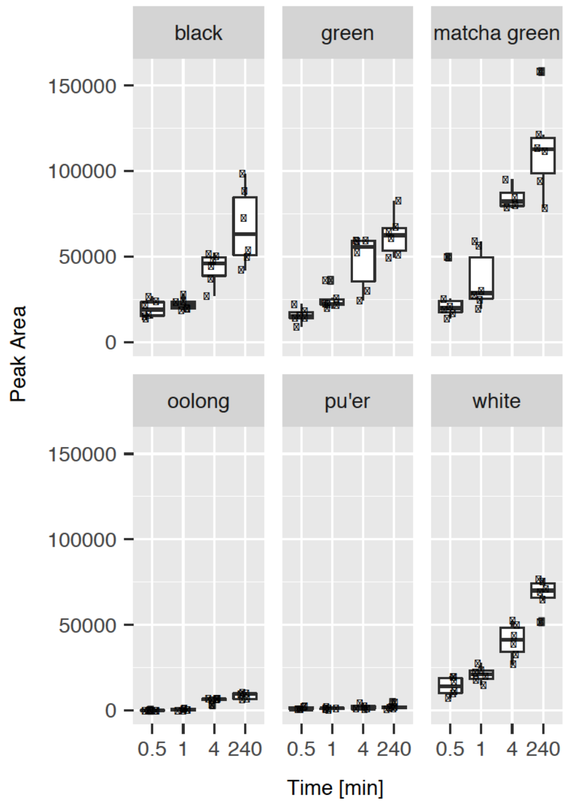
240

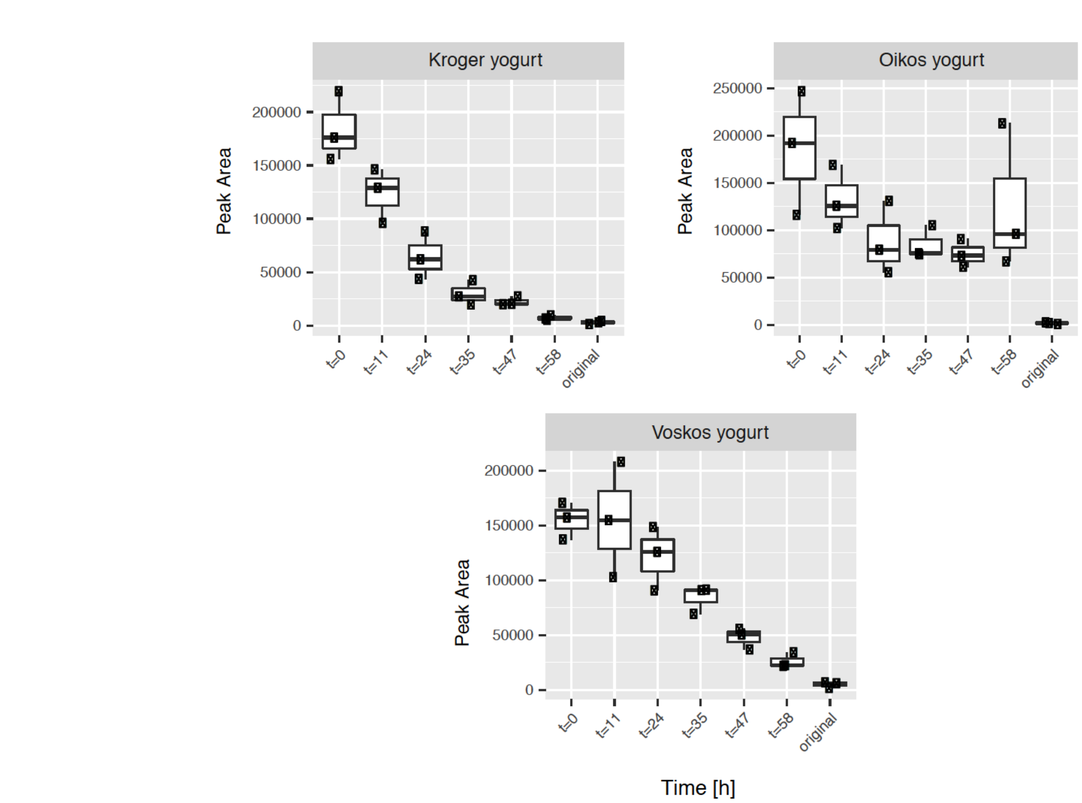
Water

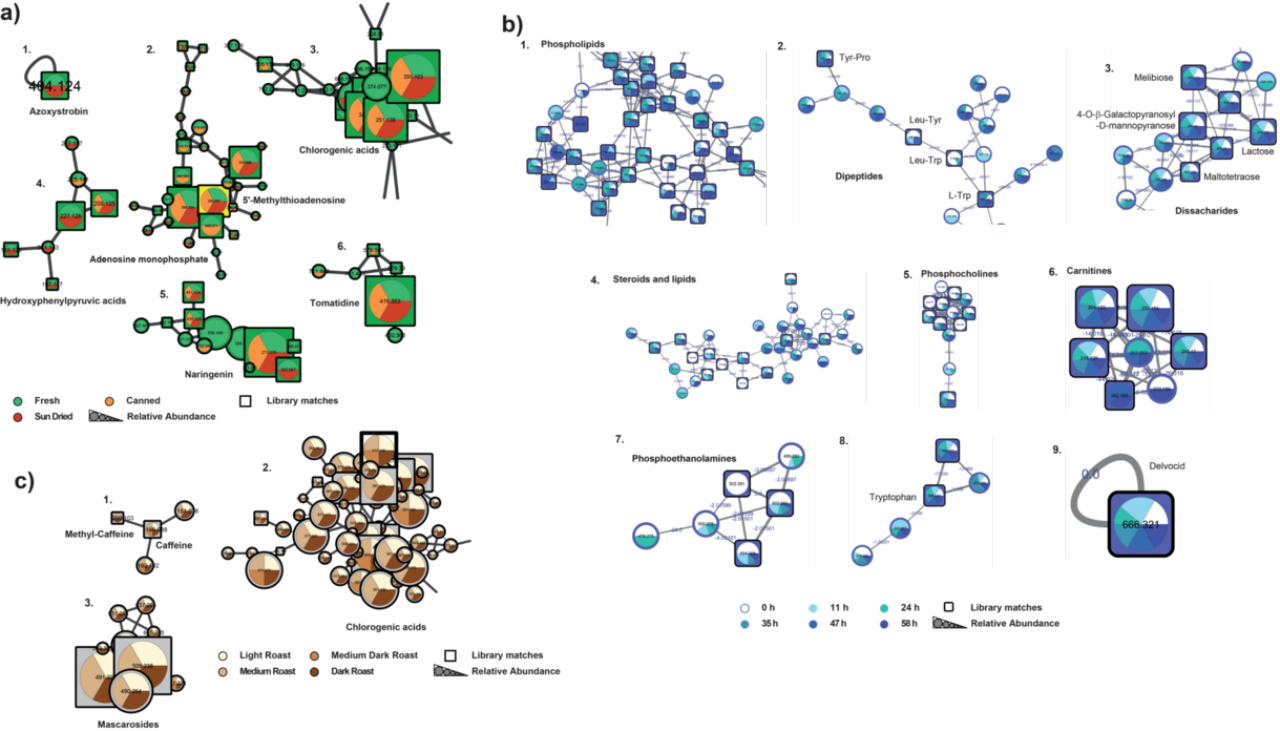
Tea leaves

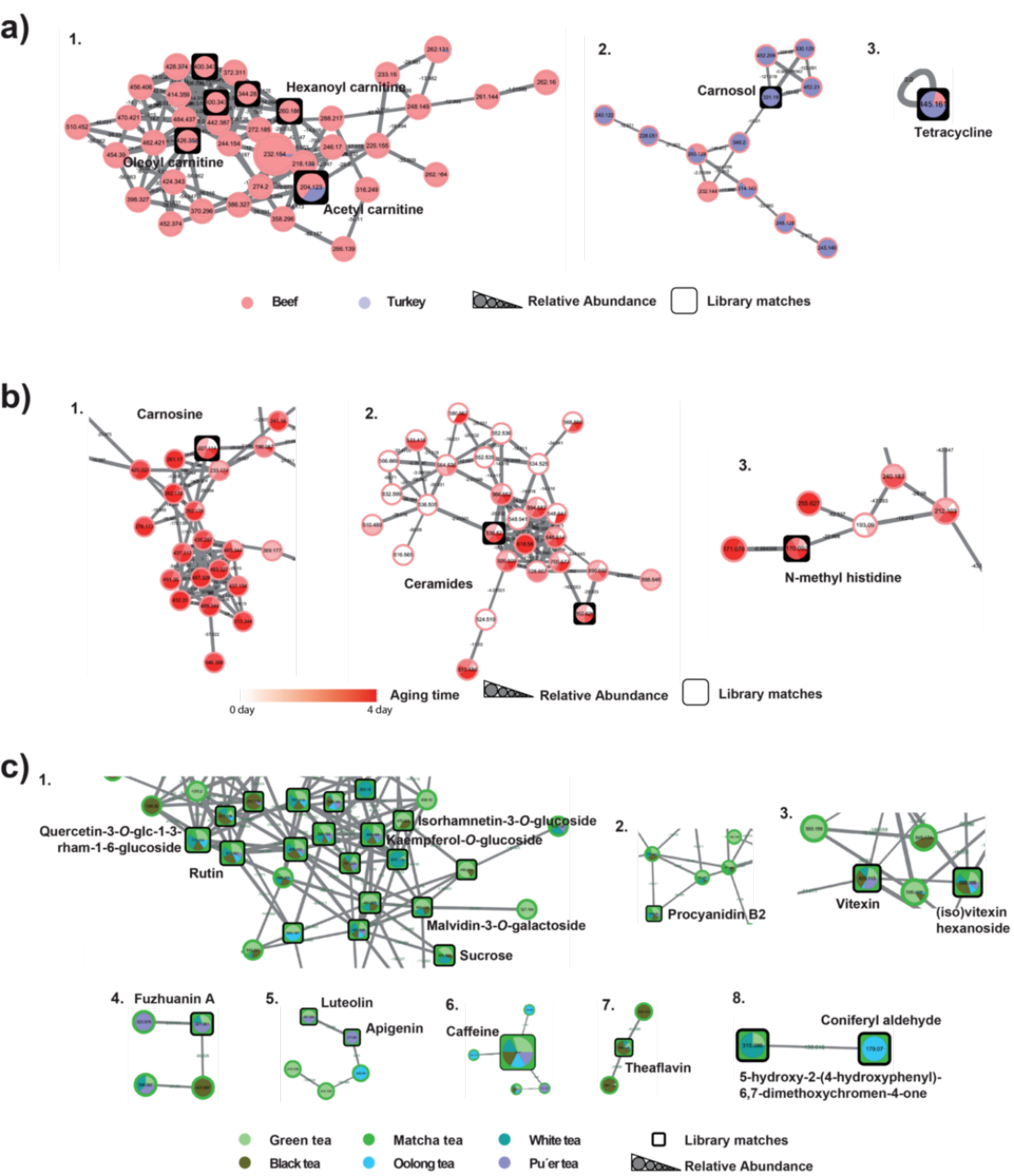

a)
black
green
matcha
oolong
pu'er
white
Time [min]
0.5 1 4 240 s
s = solid
*2
*1
*3
Row Z–Score
10
5
0
–5
–10


c)
milk
Oikos yogurt
Voskos yogurt
Kroger yogurt
Time [h]
0 11 24 35 47 58 o
o = original
*3
*2
*1
Row Z–Score
5
0
–5


b)
1. Theaflavin
black
green
matcha green
oolong
pu'er
white
Peak Area
150000
100000
50000
0
0.5 1 4 240
Time [min]


d)
1. 4–O–beta–Galactopyranosyl–
D–mannopyranose
Kroger yogurt
Oikos yogurt
Voskos yogurt
Peak Area
200000
150000
100000
50000
0
t=0
t=11
t=24
t=35
t=47
t=58
original
250000
Time [h]


a)
1.
404.124
Azoxystrobin
2.
5'-Methylthioadenosine
Adenosine monophosphate
3.
Chlorogenic acids
4.
Hydroxyphenylpyruvic acids
5.
Naringenin
6.
Tomatidine
Fresh
Canned
Library matches
Sun Dried
Relative Abundance
b)
1. Phospholipids
2.
Tyr-Pro
Leu-Tyr
Leu-Trp
L-Trp
Dipeptides
3.
Melibiose
4-O-β-Galactopyranosyl
-D-mannopyranose
Lactose
Maltotetraose
Dissaccharides
4. Steroids and lipids
5. Phosphocholines
6. Carnitines
7. Phosphoethanolamines
8.
Tryptophan
9.
Delvocid
666.321
0 h
11 h
24 h
Library matches
35 h
47 h
58 h
Relative Abundance
c)
1.
Methyl-Caffeine
Caffeine
2.
Chlorogenic acids
3.
Mascarosides
Light Roast
Medium Dark Roast
Library matches
Medium Roast
Dark Roast
Relative Abundance


a)
1.
Hexanoyl carnitine
Oleoyl carnitine
Acetyl carnitine
2.
Carnosol
3.
Tetracycline
Beef
Turkey
Relative Abundance
Library matches
b)
1.
Carnosine
2.
Ceramides
3.
N-methyl histidine
Aging time
0 day
4 day
Relative Abundance
Library matches
c)
1.
Quercetin-3-O-glc-1-3-rham-1-6-glucoside
Isorhamnetin-3-O-glucoside
Kaempferol-O-glucoside
Rutin
Malvidin-3-O-galactoside
Sucrose
2.
Procyanidin B2
3.
Vitexin
(iso)vitexin hexanoside
4. Fuzhuanin A
5.
Luteolin
Apigenin
6.
Caffeine
7.
Theaflavin
8.
Coniferyl aldehyde
5-hydroxy-2-(4-hydroxyphenyl)-6,7-dimethoxychromen-4-one
Green tea
Black tea
Matcha tea
Oolong tea
White tea
Pu´er tea
Library matches
Relative Abundance