NMR- and GC/MS-based metabolomic characterization of sunki, an unsalted fermented pickle of turnip leaves

NMR- and GC/MS-based metabolomic characterization of *sunki*, an unsalted fermented pickle of turnip leaves

Satoru Tomita[a,*], Toshihide Nakamura[a], and Sanae Okada[b]

[a] Food Research Institute, National Agriculture and Food Research Organization (NARO), 2-1-12 Kannondai, Tsukuba, Ibaraki 305-8642, Japan.

[b] Kiso Town Resource Institute, 2326-6 Fukushima, Kisomachi, Kiso, Nagano 397-8588, Japan

[*] Correspondence to S. Tomita: e-mail, satorutomita@affrc.go.jp; phone, +81-(0)29-838-8066; fax, +81-(0)29-838-7996.

Running title: Metabolomic characterization of *sunki* pickle

Abstract

This study revealed the compositional characteristics of *sunki*, a traditional, unsalted, lactic acid-fermented pickle produced using turnip leaf in Kiso district, Japan. Comprehensive compositional analysis by two metabolomic approaches based on NMR and solid-phase microextraction-GC/MS methods was used to determine its chemical composition by annotating 54 water-soluble and 62 volatile compounds. Principal component analysis showed that samples had different compositions, depending on the agricultural processing factory and production year. This variation potentially resulted from the differences in the lactic acid bacterial community produced during the spontaneous fermentation of *sunki* and in the initial nutritional composition of the turnip leaf. Partial least squares regression revealed that the acetic acid level showed a strong positive correlation with pH ($R = 0.810$), in contrast to the negative correlations of lactic acid and ethanol levels ($R = -0.533$ and -0.547). This indicated the crucial impact of acetic acid-related metabolism on acidification during *sunki* fermentation.

Keywords: Metabolomics; fermented foods; lactic acid bacteria; *Lactobacillus*; *Brassica*.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

1. Introduction

Japan has a warm and humid climate, and a wide variety of fermented foods and beverages have been developed and inherited over the centuries. Nowadays, fermented food/beverages of Japan, including soy sauce, *miso* (fermented soy bean paste), and *sake* (Japanese rice wine), are recognized worldwide (Tamang, Watanabe, & Holzapfel, 2016). *Sunki* is a unique, traditional fermented pickle, which is only produced in the Kiso district, an inland mountainous region of Nagano Prefecture, Japan. It is made by fermenting red turnip (*Brassica rapa* L.) leaves with lactic acid bacteria (LAB) (Tamang, et al., 2016). In general, salt plays a central role in pickle fermentation, as it suppresses the growth of undesirable spoilage microorganisms and promotes the release of water and nutritional compounds from the cytoplasm of vegetable plants (Henney, 2010). *Sunki* is unique in that the fermentation process occurs under unsalted conditions, in contrast to most vegetable pickles that are produced using salt. Even today, *sunki* is produced every year at the beginning of winter. In recent years, the demand for *sunki* is increasing because of the increasing awareness among the consumers about the need to reduce sodium intake and because of the health-promoting effects of LAB. This increasing demand is also making an important contribution to the growth of the rural economy.

While *sunki* is highly valuable as part of the Japanese food culture, as a biological resource, and as a product stimulating the rural economy, academic research into its fermented food properties is highly limited. However, several studies have investigated the chemical composition of *sunki*. Itabashi et al. have analyzed the levels of three organic acids (lactic acid [LacA], acetic acid [HOAc], and malic acid [MalA]), amino acids, and volatile compounds (Itabashi, 1982; Itabashi, Kawawa, & Miyao, 1990) in *sunki* samples, and provided initial information for investigating the compositional characteristics of

sunki. However, only two or three samples were analyzed in these studies, which is insufficient for establishing the chemical composition of *sunki*. It is likely that *sunki* exhibits a large compositional variety due to the following reasons: i) it is produced by spontaneous fermentation of the first-generation batch of the year and subsequently subcultured by adding a portion of this batch to the next processing batch; ii) it is produced on a small scale in homes and many small agricultural processing factories; and, iii) it is produced using three different local turnip varieties, depending on the region of the Kiso district. Therefore, to provide a compositional overview of *sunki* as a fermented food, the chemical characteristics of a larger set of samples from different places need to be investigated through a comprehensive analysis. Moreover, it is also important to survey the metabolites responsible for the pH value, because it is a practicable marker for assessing the progress of the desired fermentation of *sunki*.

In recent years, the dramatic progress in instrumental analysis has enabled the development of metabolomic approaches, which combine comprehensive compositional analysis by high-throughput analytical instruments and multivariate statistical analysis to extract the compositional differences in large datasets (metabolite profile). The metabolomic approach has been globally applied in studies on representative fermented foods and beverages containing LAB, such as yoghurt, cheese, wine, soy sauce, *miso*, *sake*, and *kimchi*, providing comprehensive information on their compositional characteristics (Hong, 2011; Lu, Hu, Miyakawa, & Tanokura, 2016; Ochi, Naito, Iwatsuki, Bamba, & Fukusaki, 2012; Park, Yoo, Seo, Lee, Na, & Son, 2016; Shiga, Yamamoto, Nakajima, Kodama, Imamura, Sato, et al., 2014; Sugimoto, Kaneko, Onuma, Sakaguchi, Mori, Abe, et al., 2012; Yoshida, Yamazaki, Ozawa, Mizukoshi, & Miyano, 2009). Moreover, these studies successfully highlighted some of the metabolites involved in the

important microbial activities of LAB, including malolactic fermentation for wine production (Hong, 2011), proteolytic activity during yoghurt production (Settachaimongkon, Nout, Antunes Fernandes, Hettinga, Vervoort, Van Hooijdonk, et al., 2014), and promotion of ripening during cheese production (Ochi, Sakai, Koishihara, Abe, Bamba, & Fukusaki, 2013).

In the present study, we applied metabolomics using two different non-targeted analytical techniques to the investigation of the chemical characteristics of *sunki* samples. The analysis was based on the comprehensive metabolite profiles of water-soluble and volatile flavor compounds obtained by nuclear magnetic resonance (NMR) spectroscopy and headspace solid-phase microextraction-gas chromatography/mass spectrometry (SPME-GC/MS), respectively. Herein, we describe the detailed chemical composition of *sunki* and provide an overview of its compositional variety, based on the data obtained using samples from eight agricultural processing factories over two years.

2. Materials and Methods

2.1. Sampling of *sunki* pickle

Sunki samples produced in November–December in 2015 and 2016 were provided from eight agricultural product-processing factories (A–H) in Kiso district, Japan, and were stored in a freezer at −20°C until use. Samples used in this study, and their pH and Na^+ concentration are listed in Table S1. The pH and Na^+ concentration were determined using a LAQUA F-72 and a LAQUAtwin Na-11 (Horiba, Kyoto, Japan), respectively. Liquid part of *sunki* pickle was used for the compositional analysis because of its advantages of high homogeneity and ease of analytical sample preparation as compared to lyophilized sample. A preliminary spectral comparison confirmed that the signal

patterns of ^{1}H NMR spectra were sufficiently consistent between the samples of liquid part and lyophilized *sunki* (Fig. S1). The samples were used for NMR and SPME-GC/MS analyses in a single analytical replicate.

2.2. NMR analysis

Water-soluble compounds of *sunki* pickle were analyzed by NMR spectroscopy. To prepare an analytical sample of pickle liquid, 140 μL of the clear supernatant was diluted with 560 μL of 125 mM potassium phosphate buffer (KPi), consisting of K_2HPO_4/KH_2PO_4 (pH 7.0) in deuterium oxide (D_2O; 99.9% D; Cambridge Isotope Laboratories, Andover MA, USA). After centrifugation, the supernatant was transferred to a 5 mm O.D. NMR tube (Norell, Landisville, NJ). Pickled leaf samples were lyophilized for one week and then ground into fine powder. Water-soluble compounds in leaf samples were extracted by suspending 10 mg of the dried powder in 700 μL of 100 mM KPi in D_2O followed by vortexing for 5 min at 25°C. After centrifugation, the supernatant was transferred to an NMR tube through a simple surgical cotton filter to remove the suspended debris. For these NMR samples, 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt (DSS-d_6; Cambridge Isotope Laboratories) was used as an internal standard at a concentration of 1.0 mM. Proton (^{1}H), carbon (^{13}C), and 2D NMR spectra were recorded on an Avance-500 spectrometer (Bruker BioSpin, Karlsruhe, Germany), using previously described acquisition parameters and conditions (Tomita, Nemoto, Matsuo, Shoji, Tanaka, Nakagawa, et al., 2015). Metabolite annotation was facilitated by analyzing the NMR spectra measured using an Avance-800 spectrometer (Bruker BioSpin) and by referring to public NMR spectral databases (SpinAssign program in the PRIMe web service, http://prime.psc.riken.jp; Human Metabolomics Database,

http://www.hmdb.ca; Biological Magnetic Resonance Data Bank, http://www.bmrb.wisc.edu).

2.3. SPME-GC/MS analysis

Volatile compounds of *sunki* pickle were investigated using a GCMS-QP2010 Ultra instrument (Shimadzu, Kyoto, Japan) equipped with an AOC-5000 autosampler (Shimadzu). A SUPELCO 50/30 μm divinylbenzene/carboxen/polydimethylsiloxane fiber (2 cm length; Sigma-Aldrich, St. Louis, MO, USA) was used for SPME. *Sunki* pickle liquid (1 mL) was transferred to a 20-mL screwcap vial and kept at 4°C during waiting time prior to measurement. The vial was preheated in the heating unit of the autosampler at 50°C for 10 min, with agitation at 250 rpm. Headspace volatile compounds were first captured by exposing the SPME fiber for 20 min and then desorbed for 5 min in an injection port operated at 250°C in splitless mode. The injected compounds were separated on an Rtx-WAX capillary column (60 m × 0.25 mm I.D. × 0.25 μm film thickness; Restek, Bellefonte, PA, USA) with a carrier gas (helium) at a flow rate of 2 mL/min. The column temperature was isothermally held at 40°C for 5 min and then raised by 5°C/min to 180°C, followed by 10°C/min to 200°C, and held for 5 min. The MS analysis was operated in the electron ionization mode with the following parameters: ionization energy of 70 eV; ion source temperature of 230°C; interface temperature of 250°C; and, scan range of 33–350 *m*/*z*. The retention index was calibrated using the alkane standard solution C_6–C_{20}. Peak annotation was performed by comparing the mass spectra and retention index with those in the NIST 02 MS library (National Institute of Standards and Technology, Gaithersburg, MD, USA). When appropriate, the MassBank spectral database (http://www.massbank.jp) was also used.

2.4. Dataset preparation for non-targeted metabolomics

154 Non-targeted, NMR-based metabolomics was performed to characterize the *sunki*
155 samples based on their water-soluble metabolite profiles. Prior to dataset preparation from
156 the ^{1}H NMR spectra, it was confirmed that there were no crucial chemical shift
157 fluctuations, which lead to significant impact on data interpretation. To prepare a dataset
158 for multivariate analysis, processed ^{1}H NMR spectra were subdivided into 0.04 ppm
159 width integral regions (buckets) in the 10.0–0.50 ppm spectral range. Twelve buckets
160 containing the residual solvent signal, ranging from 5.16 to 4.68 ppm, were excluded from
161 the analysis. To correct the difference in concentration between the pickle liquid samples,
162 the buckets were normalized to the total intensity of the NMR spectrum of each sample.
163 The generated dataset for NMR-based metabolomics comprised 226 buckets. For
164 characterization of volatile metabolite profiles by GC/MS-based metabolomics, a dataset
165 was prepared as described previously (Iijima, Iwasaki, Otagiri, Tsugawa, Sato, Otomo, et
166 al., 2016). Briefly, GC/MS raw data were processed by GCMSsolution software
167 (Shimadzu) and converted into AIA files. Baseline correction and peak alignment were
168 carried out using MetAlign software (Lommen, 2009) and the convergence of *m*/*z* ions
169 for a single compound was performed using AIoutput software (Tsugawa, Bamba,
170 Shinohara, Nishiumi, Yoshida, & Fukusaki, 2011). The variable derived from HOAc was
171 excluded from the dataset since it showed saturated peaks in all samples. The generated
172 dataset was comprised 357 peaks.

173 **2.5. Statistical analysis**

174 To analyze the difference in metabolite profiles among the *sunki* samples, principal
175 component analysis (PCA) was performed with the SIMCA software (ver. 14.0.0.1359;
176 Umetrics, Umeå, Sweden). Class separations shown in the PCA score plot was
177 statistically evaluated by Ward's hierarchical clustering analysis (HCA) using the same

178 software. For characterization of the water-soluble metabolite profiles, a multistep PCA
179 approach described previously (Nemoto, Ando, Kataoka, Arifuku, Kanazawa, Natori, et
180 al., 2007; Tomita, et al., 2015) was employed to further characterize the samples by
181 focusing on the differences in minor metabolites. The relationship between the metabolite
182 profile and pH was investigated by partial least squares (PLS) regression and the
183 generated model was evaluated by leave-one-out cross-validation. To reduce the large
184 influence derived from dominant metabolites, Pareto scaling was applied to the datasets
185 based on NMR and SPME-GC/MS analysis. Mean-centering was applied to pH values
186 for PLS analysis. The SIMCA software automatically selected the optimum number of
187 latent variables.
188
189 **3. Results**
190 **3.1. Differences in pH and Na^+ concentration and other features**
191 The pH and Na^+ concentration values of 28 *sunki* samples are listed in Table S1.
192 The Na^+ concentration was lower than 100 mg/L for all samples, reflecting the production
193 process of *sunki* as a salt-free fermented pickle. The pickle liquids differed in pH value in
194 the range of approximately 4.20–3.50, except for sample A1, which had the highest pH
195 of 4.54. The pickle liquids had no significant off-flavor and different red color intensities,
196 which originated from the red turnip pigment.
197 **3.2. Analysis of water-soluble compounds**
198 The 54 compounds annotated in this study by 1D and 2D NMR analyses are listed
199 in Table 1. As the representative data of 2D NMR analysis, metabolite annotations of the
200 ^{1}H–^{13}C heteronuclear single quantum coherence (HSQC) spectrum of sample A1 are
201 depicted in Fig. S2. The 1D NMR spectra of representative samples are depicted in Fig.

1. As shown in Table 1, 3 sugars, 4 alditols, 3 alcohols, 19 amino acids, 10 organic acids, 9 amines, and 6 other compounds were detected as the components of *sunki* pickle. The levels of glucose (Glc), fructose (Fru), sucrose (Suc), and MalA, dominant carbohydrates and organic acids in turnip leaves, were below the detection limit in most samples. Major metabolites of LAB produced via lactic fermentation of carbohydrate, including LacA, HOAc, and EtOH (Ravyts, De Vuyst, & Leroy, 2012), were detected as dominant signals. The spectra of many samples also showed high-intensity succinic acid (SucA) and mannitol signals. The signals derived from amino acids and their degradation derivatives (amines and 2-hydroxy acids) were detected with lower intensity. The relative intensity of these signals and the peak pattern characteristics across the entire spectral region differed markedly depending on the sample. For instance, samples B3 and H1 showed a higher intensity of mannitol (3.90–3.60 ppm) signals and sample G1 exhibited a noticeably weaker intensity of the HOAc (1.91 ppm) signal.

3.3. Analysis of volatile compounds

Volatile compounds detected and annotated by SPME-GC/MS analysis are listed in Table 2. Of the 62 compounds annotated, there were 15 esters, 7 ketones, 4 aldehydes, 14 alcohols, 8 nitriles, 2 sulfides, 3 isothiocyanates (ITCs), and 9 acids. Total ion chromatograms with peak annotations of representative samples are depicted in Fig. 2. Among the compounds, the HOAc peak dominated and was detected as a saturated peak. Following HOAc, two nitriles of 4-cyano-1-butene and 5-cyano-1-pentene and two ITCs of 3-butenyl ITC and 4-pentenyl ITC exhibited intense signals. These four volatiles were dominant in most samples, although several samples, especially sample G1, showed markedly lower intensity peaks for the two ITCs (Fig. 2). Additionally, EtOH and ethyl acetate were detected as the secondary major peaks. Their intensities appeared to be

highly sample-dependent; the peaks were strongly detected in samples A1, B1, B3, and G1, but observed with low intensity in samples C4 and E2 (Fig. 2). Other compounds were observed as minor peaks on the chromatograms and the peak pattern varied among the samples.

3.4. NMR-based metabolomic characterization of *sunki* samples

As a first step to characterize *sunki* samples based on the profile of water-soluble compounds, PCA was carried out using the dataset containing all 226 buckets generated from the ^{1}H NMR spectra. The score and loading plots are shown in Fig. 3a. The first and second principal components (PC1 and PC2) explained 41.0% and 25.2% of the total variance, respectively. In the score plot, the samples were divided by Ward's HCA into two classes as shown in Fig. 3a. The separation did not simply associate with the factory or production year and each class comprised samples from different factories and years. Specifically, the samples from factories A, B, D, and F were separated into these two classes in association with the production year, whereas those from factories C, G, and E belonged within the same class (Fig. 3a). The loading plot showed that PC1 was primarily explained by the variables containing signals derived from LacA, HOAc, SucA, EtOH, methanol, and mannitol. For the PC2 axis, the level of HOAc showed significant contribution and strongly characterized sample A1. Of these, it was statistically confirmed that the variables of LacA (1.30 and 1.34 ppm) and mannitol (3.78 and 3.86 ppm) were significantly different between the two classes ($p < 0.01$). This initial PCA revealed the *sunki* sample characteristics contributed by the dominant metabolites produced by LAB.

Next, to observe the contribution of minor metabolites, multistep PCA was performed by excluding the variables containing signals of dominant metabolites highlighted above. Variables containing those of residual sugars (Glc and Fru) were also

successively excluded as they substantially characterized only samples C4 and C5. After excluding these 27 variables, the generated PCA model provided an altered feature space, as shown in Fig. 3b. PC1 and PC2 explained 34.9% and 20.4% of the total variance, respectively. In the score plot, PC1 particularly highlighted some samples produced in 2015 (A1, A2, B1, B2, C1, and F1), and the class separation of these samples were confirmed by HCA (Fig. 3b). The loading plot indicated that the separation of this class was primarily explained by the difference in signal intensities derived from alanine (Ala), branched-chain amino acids (BCAAs, including valine, leucine, and isoleucine), γ-aminobutyric acid (GABA), putrescine (Put), and cadaverine (Cad). Furthermore, signal intensities of Ala, BCAAs, Put&Cad (1.46, 0.98, and 1.70, respectively) were statistically evaluated to be significantly greater in the highlighted class ($p < 0.01$). Besides, PC2 mainly explained the variety within the classes by the opposed contributions of GABA and glutamic acid (Glu).

3.5. GC/MS-based metabolomic characterization of *sunki* samples

Subsequently, the differences in the volatile compound profiles among *sunki* samples were investigated by PCA. The data from PCA and HCA showed a clear separation between the samples produced in 2015 and 2016 (Fig. 3c). The separation was along the PC1 axis (26.7% of the total variance). The class of samples produced in 2016 was mainly explained by the higher levels of ITCs (3-butenyl ITC and 4-pentenyl ITC), whereas the class of the samples in 2015 was explained by greater intensities of ethyl esters (ethyl acetate and ethyl lactate). The significant differences in the intensities of 3-butenyl ITC and ethyl acetate were statistically evaluated ($p < 0.01$). It was also observed that the intensity of 4-pentenyl ITC differed between the separated classes ($p = 0.07$). The EtOH level, which substantially explained PC2 (13.8% of the total variance) in

cooperation with EtOAc, contributed to the variety within the classes rather than the separation by the production year (Fig. 3c). The PCA data indicated that the difference in the volatile compound profile of *sunki* was primarily associated with the production year. Subsequently, to compare the compositional differences within the same production year, successive characterization was performed by PCA (Fig. S3). The PCA data showed that ITCs, ethyl esters, and EtOH also contributed to the compositional variety among the samples of the same production year. In addition, contributions of other volatiles were observed, e.g., nitriles (benzenepropanenitrile, 4-cyano-1-butene, and 5-cyano-1-pentene), 3-hexene-1-ol, and 3-hexenyl acetate.

3.6. Multivariate analysis for correlation between pH value and metabolite profile

PLS regression was carried out to investigate the correlation between pH and water-soluble metabolite profile of the liquid samples. The generated PLS model comprised two latent variables (first and second PLS components), which explained 27.5% and 37.3% of the total variance, respectively. The cumulative determination coefficient (R^2) and cumulative cross-validation determination coefficient (Q^2) were 0.937 and 0.866, respectively. The variance importance for projection (VIP) scores showed that HOAc was the most important factor for the PLS model and contributed to a substantial extent compared to other dominant metabolites, including LacA, EtOH, mannitol, and SucA (Fig. 4a). Intriguingly, the integral of HOAc exhibited a high positive correlation with pH, providing an R value of 0.810, whereas those of LacA and EtOH indicated negative correlations with pH, with R values of −0.533 and −0.547, respectively (Fig. 4b). For the volatile flavor compound profiles, a PLS model comprising three components was obtained, which afforded cumulative R^2 and Q^2 values of 0.983 and 0.878, respectively. The first two components explained 27.4% and 14.5% of the total variance, respectively.

298 Among the volatile compounds, EtOH and ethyl acetate had the highest VIP scores of 7.9
299 and 6.9, respectively. The signal intensities of these compounds exhibited significant
300 negative correlations with pH ($p < 0.05$), providing R values of −0.601 and −0.767,
301 respectively.
302
303 **4. Discussion**
304 The present study determined the chemical composition of *sunki* and provided an
305 overview of its compositional variety across eight factories and two years. For water-
306 soluble compounds, the presences of LacA, HOAc, MalA, and amino acids are in
307 agreement with previous reports on the chemical composition of *sunki* (Itabashi, 1982;
308 Itabashi, et al., 1990). LacA, the most dominant metabolite in all pickle liquids, is a major
309 product of LAB through lactic fermentation of carbohydrates. It is suggested that EtOH
310 and HOAc in *sunki* are also produced by LAB via heterolactic fermentation pathway,
311 which catabolizes Glc not only to LacA but also to CO_2, EtOH, and HOAc (Ravyts, et al.,
312 2012). In contrast, homo-fermentative LAB produce two moles of LacA from one mole
313 of Glc. The dominant carbohydrates of turnip leaves (Glc, Fru, and Suc) were
314 undetectable in most samples, indicating the complete consumption of these compounds
315 during the fermentation of *sunki*. Therefore, as with other lactic-fermented pickles of
316 *Brassica* vegetables such as *sauerkraut*, *suan-cai*, and *gundruk* (Plengvidhya, Breidt, Lu,
317 & Fleming, 2007; Tamang & Tamang, 2010; Yang, Zou, Qu, Zhang, Liu, Wu, et al., 2014),
318 substantial accumulation of LacA by LAB appears to be essential for *sunki* production.
319 MalA (followed by citric acid) is a dominant organic acid in the leaves of endemic turnip
320 varieties used for *sunki* production in the Kiso district (Fig. S4). The fact that MalA was
321 also undetectable in most samples indicates its active consumption during in *sunki*

322 production. Amino acids in *sunki* would largely be derived from turnip leaves, since LAB
323 species, in general, are auxotrophic for multiple amino acids (Wegkamp, Teusink, De Vos,
324 & Smid, 2010). Proteolytic activity of LAB during fermentation could also be a cause of
325 the presence of amino acids.

326 Moreover, in addition to these known components of *sunki*, the following
327 metabolites were annotated in the present study. SucA was detected in most samples and
328 it would be mainly produced by fermentation, as SucA production by LAB and its
329 promoting effect by the presences of MalA and citric acid have been reported (Kaneuchi,
330 Seki, & Komagata, 1988). Mannitol is known to be a compound directly converted from
331 Fru by mannitol dehydrogenase of LAB (Wisselink, Weusthuis, Eggink, Hugenholtz, &
332 Grobben, 2002). Amines and 2-hydroxy acids were annotated as minor components of
333 *sunki*. These compounds have been detected in lactic-fermented pickles (Kalač, Špička,
334 KříŽek, & Pelikánová, 2000; Wu, Zheng, Huang, & Zhou, 2014), and reported as amino
335 acid degradation products resulting from decarboxylation and transamination (Shalaby,
336 1996; Smid & Kleerebezem, 2014). Acetoin and 2,3-butanediol were annotated in several
337 samples and these are generally produced from pyruvic acid via 2,3-butanedione.
338 Ascorbic acid and methiin (also known as *S*-methylcysteine sulfoxide) were probably
339 derived from the plant material since the former is present in abundance in the turnip leaf
340 (National Nutrient Database: https://ndb.nal.usda.gov/ndb/) and the latter is a well-known
341 sulfur compound of *Brassica* vegetables.

342 Regarding the composition of volatile flavor compounds, many of the annotated
343 peaks appear to originate from the turnip leaf rather than by microbial action during
344 fermentation. ITCs and nitriles are known to be unique volatiles, which characterize the
345 flavor of *Brassica* vegetables and are produced through the degradation of glucosinolates

346 by the catalytic activity of myrosinase (Rask, Andreasson, Ekbom, Eriksson, Pontoppidan,
347 & Meijer, 2000). Of the four dominant volatiles detected in all samples in this study, it
348 was suggested that 3-butenyl ITC and 4-cyano-1-butene were derived from gluconapin,
349 and 4-pentenyl ITC and 5-cyano-1-pentene were derived from glucobrassicanapin. These
350 assignments are in good agreement with the data from a previous study that described
351 gluconapin and glucobrassicanapin as dominant glucosinolates in the turnip leaf grown
352 in Japan (Osada & Aoyagi, 2014). Itabashi et al. have performed GC/MS analysis of three
353 *sunki* samples to determine their volatile composition (Itabashi, et al., 1990). They
354 detected four intense peaks and estimated three of them as corresponding to ITCs and
355 nitriles. The four dominant ITCs and nitriles annotated in the present study suggest to
356 correspond to the four intense peaks in the previous study. Compared to previous reports,
357 it was found that *sunki* differed from other fermented pickles (*sauerkraut* and *suan-cai*)
358 in terms of the ITC and nitrile profiles (Palani, Harbaum-Piayda, Meske, Keppler,
359 Bockelmann, Heller, et al., 2016; Wu, Yu, Liu, Meng, Wang, Xue, et al., 2015). This
360 probably indicates that ITC and nitrile profiles reflect the native composition of
361 glucosinolates in *Brassica* crops utilized for producing each pickle. In contrast with the
362 composition of the dominant glucosinolate-derived compounds, the composition of the
363 volatiles resulting from microbial action were basically similar between *sunki* and those
364 of other fermented pickles. EtOH, HOAc, formic acid, acetoin, 2,3-butanedione, ethyl
365 acetate, and fatty acids are representative compounds in lactate-fermented foods.

366 Non-targeted metabolomic characterization revealed the differences in the
367 metabolite profiles of the *sunki* samples. For characterization of the water-soluble
368 metabolite profiles, compounds primarily contributing to PC1 (LacA, EtOH, HOAc, and
369 mannitol) were correlated with the main products of homo- and hetero-lactic fermentative

pathways. Although mannitol is not a component of lactate fermentation, Wisselink et al. have described that hetero-fermentative LAB produce mannitol in large amounts to maintain the intracellular redox balance (Wisselink, et al., 2002). Endo et al. have reported that homo-fermentative *L. plantarum* and *L. delbrueckii* and hetero-fermentative *L. fermentum* were dominant during *sunki* fermentation (Endo, Mizuno, & Okada, 2008). Therefore, the PCA data shown in Fig. 3a might indicate that the separation along the PC1 axis reflects the variation in the dominance of these lactobacilli during fermentation. Similarly, the year-to-year separation for the samples from factories A, B, D, and F suggests the possibility that bacterial community contributing to *sunki* fermentation and the resulting food composition vary even in the same processing factory, depending on the production year.

HOAc is a product of heterolactic fermentation as described above. However, its signal level was not associated with that of other hetero-fermentative products (EtOH and mannitol). HOAc is rather substantially responsible for PC2, while showing inverted contributions against LacA and EtOH, resulting in the secondary separation of samples along the PC2 axis. This observation regarding HOAc has not been reported in previous studies on *sunki* or, to the best of our knowledge, on spontaneous lactate-fermented pickles including *sauerkraut*, *suan-cai*, and *gundruk*. Although metabolism leading to HOAc accumulation remains unclear from the data obtained in this study, it is notable that the HOAc signal intensity showed a higher correlation with pH than that of LacA, which is supposed to be the primal factor responsible for a decrease in pH (Fig. 4b). It is also intriguing that the HOAc content was positively correlated with pH, in contrast to the LacA and EtOH levels that showed negative correlations with pH. A sufficient decrease in pH is the most important marker for the quality of *sunki*, and is essential for

assessing fermentation progress. Nevertheless, excessive acidification is not desirable, as it creates a strong sour taste. Clarification of the mechanism of HOAc accumulation and the impact on pH in terms of microbial and metabolic aspects would help prevent poor/excessive acidification in the production of *sunki* and other lactic-fermented pickles.

A multistep PCA approach excluding the dominant water-soluble compounds principally highlighted the differences among the samples associated with factories and production years (Fig. 3b), but not with the class separation between homo- and hetero-fermentative types shown in the first PCA (Fig. 3a). The differences observed in the multistep PCA appeared to be the result of variation in the microbial metabolic activity, since many of the compounds responsible for PC1 and PC2 are decarboxylation and transamination products of amino acids. The highlighted compounds, Ala, Put, Cad, and GABA, are converted from aspartic acid, ornithine, lysine, and Glu, respectively, by the catalytic action of amino acid decarboxylases of LABs (Kalač, et al., 2000). The opposed contributions of Glu and GABA shown in the loading plot of Fig. 3b probably reflected the relationship of their interconversion. Hence, it can be estimated that Glu decarboxylase-active LABs were dominant in the fermentation of the *sunki* samples characterized by the greater GABA level. BCAAs and Glu were suggested to originate from the turnip leaf, as well as being possibly accumulated by the proteolytic activity of LAB. Amino acids and their degradation products can impact the fermented food quality, such as sensory properties (e.g., *umami* taste of Glu and aroma of phenyllactic acid), and beneficial effects associated with GABA. With regard to food safety, biogenic amines should be maintained at low levels (Kalač, et al., 2000); the levels of tyramine and histamine detected in several *sunki* samples were estimated to be lower than 17.0 and 11.3 mg/L, respectively, by relative signal integrals to internal standard. These potential

418 metabolites detected in *sunki* and their metabolism by LAB could be utilized for
419 improving and controlling the product quality.

420 In terms of the volatile compounds of *sunki*, an evident separation between the
421 samples produced in 2015 and 2016 was observed. Successive PCA within the same
422 production year also revealed differences among the factories. These results suggest that
423 the volatile compound profile of *sunki* differs depending on the production year and
424 factory, based on the main contributions of different levels of ITCs, ethyl esters, nitriles,
425 3-hexene-1-al, and 3-hexenyl acetate. With respect to glucosinolate-derived ITCs and
426 nitriles, in the salted pickle production of *Brassica* vegetables in Japan, it has been
427 reported that the levels of these compounds depend on various processing conditions such
428 as the initial glucosinolate profile of plant material, blanching and residual myrosinase
429 activity, pH change during pickling, and the presence of antioxidative constituents (Kato,
430 Imayoshi, Iwabuchi, & Shimomura, 2011; Uda, Yabe, Sueki, Suzuki, & Maeda, 1991).
431 These factors possibly also have an impact on the ITC and nitrile levels of *sunki*. Besides,
432 as the bacterial myrosinase activity has been reported in *Lactobacillus agilis* (Palop,
433 Smiths, & Tenbrink, 1995), variety in the bacterial community could affect the ITC and
434 nitrile formation during *sunki* fermentation. With respect to the fermentation-derived
435 compounds, EtOH, ethyl acetate, and ethyl lactate were highlighted by GC/MS-based
436 PCA. These two esters provide a fruity aroma and have been detected as dominant esters
437 in *suan-cai* (Wu, et al., 2015). The distinct level of ethyl lactate might relate to the initial
438 MalA content of the turnip leaf as Pozo-Bayon et al. described that their relationship in
439 malolactic fermentation of wine (Pozo-Bayon, Alegria, Polo, Tenorio, Martin-Alvarez,
440 De La Banda, et al., 2005). In addition, the distinct levels of the following compounds
441 highlighted by PCA might impact the aromatic character of *sunki*. 3-Hexene-1-ol (also

known as leaf alcohol) and its acetic ester of 3-hexenyl acetate are well known to be responsible for the fresh, green grass-like odor of vegetables. Benzenepropanenitrile is also an odor compound derived from the degradation of gluconasturtiin, which is a glucosinolate known to be present in the turnip leaf (Osada, et al., 2014). Although fatty acids, which are typical off-flavor volatiles produced by the degradation of lipids during fermentation, were detected as minor peaks, they did not contribute significantly to the characteristics among the samples, in agreement with the observation that no significant off-flavor was recognized in all samples.

We showed that the chemical composition of *sunki* varied among the samples depending on the factory and production year, and correlated with the pH values. Differences in the bacterial community during *sunki* fermentation and initial composition of raw materials are most likely to have a direct impact on the metabolite profile and the resulting sensory qualities and potential health promoting effects. Further studies focusing on the triangular relationship between metabolite profile–microbiota–sensory qualities will facilitate the development of starter cultures or fermentation-controlling technology of *sunki*, leading to an improved overall quality of fermented pickles made from various vegetables.

5. Acknowledgements

We thank Mr. Kiyotsugu Shimizu (Kiso Town) and managers of the agricultural processing factories in Kiso district for kindly providing *sunki* samples. NMR analyses were carried out with the technical support of the Advanced Analysis Center at National Agriculture and Food Research Organization (Tsukuba, Japan).

6. Conflict of Interest Statement

The authors have declared no conflict of interest.

7. References

Endo, A., Mizuno, H., & Okada, S. (2008). Monitoring the bacterial community during fermentation of sunki, an unsalted, fermented vegetable traditional to the Kiso area of Japan. *Letters in Applied Microbiology, 47*(3), 221-226.

Henney, J. E. (2010). *Strategies to reduce sodium intake in the United States*. Washington (DC): National Academies Press, (Chapter 4).

Hong, Y. S. (2011). NMR-based metabolomics in wine science. *Magnetic Resonance in Chemistry, 49 Suppl 1*(S1), S13-21.

Iijima, Y., Iwasaki, Y., Otagiri, Y., Tsugawa, H., Sato, T., Otomo, H., Sekine, Y., & Obata, A. (2016). Flavor characteristics of the juices from fresh market tomatoes differentiated from those from processing tomatoes by combined analysis of volatile profiles with sensory evaluation. *Bioscience Biotechnology and Biochemistry, 80*(12), 2401-2411.

Itabashi, M. (1982). Studies on the Japanese pickles sunki (I) Dietetic components of the sunki. *Science of Cookery, 15*(4), 226-228. (in Japanese)

Itabashi, M., Kawawa, Y., & Miyao, S. (1990). Flavor components of *sunki* pickles. *Nippon Shokuhin Kagaku Kogaku Kaishi, 37*(1), 15-19. (in Japanese)

Kalač, P., Špička, J., Křížek, M., & Pelikánová, T. (2000). The effects of lactic acid bacteria inoculants on biogenic amines formation in sauerkraut. *Food Chemistry, 70*(3), 355-359.

Kaneuchi, C., Seki, M., & Komagata, K. (1988). Production of succinic acid from citric-acid and related acids by *Lactobacillus* strains. *Applied and Environmental Microbiology, 54*(12), 3053-3056.

Kato, M., Imayoshi, Y., Iwabuchi, H., & Shimomura, K. (2011). Kinetic changes in glucosinolate-derived volatiles by heat-treatment and myrosinase activity in Nakajimana (*Brassica rapa* L. cv. *nakajimana*). *Journal of Agricultural and Food Chemistry, 59*(20), 11034-11039.

Lommen, A. (2009). MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Analytical Chemistry, 81*(8), 3079-3086.

Lu, Y., Hu, F., Miyakawa, T., & Tanokura, M. (2016). Complex mixture analysis of organic compounds in yogurt by NMR spectroscopy. *Metabolites, 6*(2), 19.

Nemoto, T., Ando, I., Kataoka, T., Arifuku, K., Kanazawa, K., Natori, Y., & Fujiwara, M. (2007). NMR metabolic profiling combined with two-step principal component analysis for toxin-induced diabetes model rat using urine. *Journal of Toxicological Sciences, 32*(4), 429-435.

Ochi, H., Naito, H., Iwatsuki, K., Bamba, T., & Fukusaki, E. (2012). Metabolomics-based component profiling of hard and semi-hard natural cheeses with gas chromatography/time-of-flight-mass spectrometry, and its application to sensory predictive modeling. *Journal of Bioscience and Bioengineering, 113*(6), 751-758.

509 Ochi, H., Sakai, Y., Koishihara, H., Abe, F., Bamba, T., & Fukusaki, E. (2013).
510 Monitoring the ripening process of Cheddar cheese based on hydrophilic
511 component profiling using gas chromatography-mass spectrometry. *Journal of*
512 *Dairy Science, 96*(12), 7427-7441.
513 Osada, S., & Aoyagi, Y. (2014). Level of glucosinolates in Brassicaceae vegetables
514 harvested during autumn and winter in Japan. *Journal for the Integrated Study of*
515 *Dietary Habits, 25*(2), 121-130. (in Japanese)
516 Palani, K., Harbaum-Piayda, B., Meske, D., Keppler, J. K., Bockelmann, W., Heller, K.
517 J., & Schwarz, K. (2016). Influence of fermentation on glucosinolates and
518 glucobrassicin degradation products in sauerkraut. *Food Chemistry, 190*, 755-762.
519 Palop, M. L., Smiths, J. P., & Tenbrink, B. (1995). Degradation of sinigrin by
520 *Lactobacillus agilis* strain R16. *International Journal of Food Microbiology,*
521 *26*(2), 219-229.
522 Park, S. E., Yoo, S. A., Seo, S. H., Lee, K. I., Na, C. S., & Son, H. S. (2016). GC-MS
523 based metabolomics approach of Kimchi for the understanding of *Lactobacillus*
524 *plantarum* fermentation characteristics. *LWT-Food Science and Technology, 68*,
525 313-321.
526 Plengvidhya, V., Breidt, F., Lu, Z., & Fleming, H. P. (2007). DNA fingerprinting of lactic
527 acid bacteria in sauerkraut fermentations. *Applied and Environmental*
528 *Microbiology, 73*(23), 7697-7702.
529 Pozo-Bayon, M. A., Alegria, E. G., Polo, M. C., Tenorio, C., Martin-Alvarez, P. J., De
530 La Banda, M. T. C., Ruiz-Larrea, F., & Moreno-Arribas, M. V. (2005). Wine
531 volatile and amino acid composition after malolactic fermentation: Effect of
532 *Oenococcus oeni* and *Lactobacillus plantarum* starter cultures. *Journal of*
533 *Agricultural and Food Chemistry, 53*(22), 8729-8735.
534 Rask, L., Andreasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B., & Meijer, J. (2000).
535 Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant*
536 *Molecular Biology, 42*(1), 93-113.
537 Ravyts, F., De Vuyst, L., & Leroy, F. (2012). Bacterial diversity and functionalities in
538 food fermentations. *Engineering in Life Sciences, 12*(4), 356-367.
539 Settachaimongkon, S., Nout, M. J., Antunes Fernandes, E. C., Hettinga, K. A., Vervoort,
540 J. M., van Hooijdonk, T. C., Zwietering, M. H., Smid, E. J., & van Valenberg, H.
541 J. (2014). Influence of different proteolytic strains of *Streptococcus thermophilus*
542 in co-culture with *Lactobacillus delbrueckii* subsp. *bulgaricus* on the metabolite
543 profile of set-yoghurt. *International Journal of Food Microbiology, 177*, 29-36.
544 Shalaby, A. R. (1996). Significance of biogenic amines to food safety and human health.
545 *Food Research International, 29*(7), 675-690.
546 Shiga, K., Yamamoto, S., Nakajima, A., Kodama, Y., Imamura, M., Sato, T., Uchida, R.,
547 Obata, A., Bamba, T., & Fukusaki, E. (2014). Metabolic profiling approach to
548 explore compounds related to the umami intensity of soy sauce. *Journal of*
549 *Agricultural and Food Chemistry, 62*(29), 7317-7322.
550 Smid, E. J., & Kleerebezem, M. (2014). Production of aroma compounds in lactic
551 fermentations. *Annu Rev Food Sci Technol, 5*, 313-326.
552 Sugimoto, M., Kaneko, M., Onuma, H., Sakaguchi, Y., Mori, M., Abe, S., Soga, T., &
553 Tomita, M. (2012). Changes in the charged metabolite and sugar profiles of
554 pasteurized and unpasteurized Japanese sake with storage. *Journal of Agricultural*
555 *and Food Chemistry, 60*(10), 2586-2593.

556 Tamang, B., & Tamang, J. P. (2010). *In situ* fermentation dynamics during production of
557 *gundruk* and *khalpi*, ethnic fermented vegetable products of the Himalayas. *Indian*
558 *journal of microbiology, 50*, S93-S98.
559 Tamang, J. P., Watanabe, K., & Holzapfel, W. H. (2016). Review: diversity of
560 microorganisms in global fermented foods and beverages. *Frontiers in*
561 *Microbiology, 7*, 377.
562 Tomita, S., Nemoto, T., Matsuo, Y., Shoji, T., Tanaka, F., Nakagawa, H., Ono, H.,
563 Kikuchi, J., Ohnishi-Kameyama, M., & Sekiyama, Y. (2015). A NMR-based,
564 non-targeted multistep metabolic profiling revealed L-rhamnitol as a metabolite
565 that characterised apples from different geographic origins. *Food Chemistry, 174*,
566 163-172.
567 Tsugawa, H., Bamba, T., Shinohara, M., Nishiumi, S., Yoshida, M., & Fukusaki, E.
568 (2011). Practical non-targeted gas chromatography/mass spectrometry-based
569 metabolomics platform for metabolic phenotype analysis. *Journal of Bioscience*
570 *and Bioengineering, 112*(3), 292-298.
571 Uda, Y., Yabe, E., Sueki, K., Suzuki, K., & Maeda, Y. (1991). Effects of sodium
572 ascorbate on pH, apparent color, volatile isothiocyanates and their related volatile
573 components of lightly pickled radish (*Raphanus sativus* L) roots and takana
574 (*Brassica juncea* Czern et Coss) leaves. *Nippon Shokuhin Kagaku Kogaku Kaishi,*
575 *38*(1), 55-61. (in Japanese)
576 Wegkamp, A., Teusink, B., de Vos, W. M., & Smid, E. J. (2010). Development of a
577 minimal growth medium for *Lactobacillus plantarum*. *Letters in Applied*
578 *Microbiology, 50*(1), 57-64.
579 Wisselink, H. W., Weusthuis, R. A., Eggink, G., Hugenholtz, J., & Grobben, G. J. (2002).
580 Mannitol production by lactic acid bacteria: a review. *International Dairy Journal,*
581 *12*(2-3), 151-161.
582 Wu, C. D., Zheng, J., Huang, J., & Zhou, R. Q. (2014). Reduced nitrite and biogenic
583 amine concentrations and improved flavor components of Chinese sauerkraut via
584 co-culture of *Lactobacillus plantarum* and *Zygosaccharomyces rouxii*. *Annals of*
585 *Microbiology, 64*(2), 847-857.
586 Wu, R. N., Yu, M. L., Liu, X. Y., Meng, L. S., Wang, Q. Q., Xue, Y. T., Wu, J. R., &
587 Yue, X. Q. (2015). Changes in flavour and microbial diversity during natural
588 fermentation of *suan-cai*, a traditional food made in Northeast China.
589 *International Journal of Food Microbiology, 211*, 23-31.
590 Yang, H. Y., Zou, H. F., Qu, C., Zhang, L. Q., Liu, T., Wu, H., & Li, Y. H. (2014).
591 Dominant microorganisms during the spontaneous fermentation of suan cai, a
592 Chinese fermented vegetable. *Food Science and Technology Research, 20*(5),
593 915-926.
594 Yoshida, H., Yamazaki, J., Ozawa, S., Mizukoshi, T., & Miyano, H. (2009). Advantage
595 of LC-MS metabolomics methodology targeting hydrophilic compounds in the
596 studies of fermented food samples. *Journal of Agricultural and Food Chemistry,*
597 *57*(4), 1119-1126.

598 **Figure Captions**

599 **Fig. 1. Representative ^{1}H NMR spectra and metabolite annotations of *sunki* pickle.**
600 Spectra of the samples A1, A4, B3, G1, and H1 are displayed in the different spectral
601 ranges: 8.50–5.00 ppm (top), 4.80–0.80 ppm (middle), enlarged view of 4.80–0.80 ppm
602 (bottom). Signal intensity was normalized by the signal of internal standard (DSS) at 0.00
603 ppm and adjusted for each range displayed. Numerical labels indicate the signals of
604 annotated metabolites listed in Table 1. The label U represents the signal of unannotated
605 metabolites.

606 **Fig. 2. Representative GC/MS chromatograms and metabolite annotations of *sunki***
607 **pickle.** Raw total ion chromatograms (top) of the samples A1, B1, B3, C4, E2, and G1,
608 and their enlarged views (bottom) are depicted. Numerical labels indicate the peaks of
609 annotated metabolites listed in Table 2. Asterisk represents the peaks observed in blank
610 measurements.

611 **Fig. 3. Non-targeted metabolomic characterization of *sunki* samples by PCA.** PC1–
612 PC2 planes of score (left) and loading (right) plots obtained from (a) NMR-based PCA,
613 (b) NMR-based multistep PCA, and (c) GC/MS-based PCA. In the score plots, the
614 samples are represented by different symbols and colors as shown in the legend. Broken
615 line represents class separation indicated by HCA. In the loading plots, numerical labels
616 represent chemical shift (ppm) for panels a and b, and retention time (min) for panel c.
617 The representative variables are also labeled with metabolite names. Abbreviations: LacA,
618 lactic acid; HOAc, acetic acid; SucA, succinic acid; MeOH, methanol; EtOH, ethanol;
619 Ala, alanine; BCAAs, blanched chain amino acids; Phe, phenylalanine; PLA, phenyllactic
620 acid; Tyr, tyrosine; Tym, tyramine; GABA, γ-aminobutyric acid; Glu, glutamic acid; Gln,
621 glutamine; Glp, pyroglutamic acid, EtOAc, ethyl acetate; ITC, isothiocyanate.

622 **Fig. 4. Results of correlation analysis between metabolite profile and pH value. (**a)
623 VIP scores obtained by NMR-based PLS regression. (b) Correlation of pH value with
624 selected raw integral values relative to internal standard (DSS).
625
626 **Supplementary Material Descriptions**
627 **Table S1. List of *sunki* samples used in this study.**
628 **Fig. S1. Compatibility of ^{1}H NMR spectra between the samples of liquid part and**
629 **lyophilized leaves of *sunki*.** Overlapping spectra of the extract from lyophilized *sunki*
630 (blue) over that from liquid part (red) are displayed. The spectral range 4.50–0.80 ppm
631 (top) and the same range with enlarged intensity (bottom) are shown.
632 **Fig. S2. Representative HSQC spectrum of pickle liquid.** The spectrum was measured
633 on an 800 MHz NMR instrument. Two divided spectral ranges are depicted in panels a
634 and b. Numerical labels indicate the signals of annotated metabolites listed in Table 1.
635 The label U represents the signal of unannotated metabolites.
636 **Fig. S3. GC/MS-based PCA of *sunki* samples collected in 2015 and 2016.** The PC1–
637 PC2 planes of score (left) and loading (right) plots are displayed. In the score plots, the
638 symbols are color-coded according to the agricultural processing factory and the samples
639 collected in 2015 and 2016 are represented as filled circles and open boxes, respectively.
640 In the loading plots, numerical labels represent retention time (min). The representative
641 variables are also labeled with metabolite names. Metabolite abbreviations: EtOH,
642 ethanol; EtOAc, ethyl acetate; ITC, isothiocyanate.
643 **Fig. S4. ^{1}H NMR spectrum of the extract from lyophilized turnip leaf.** The spectral
644 range 5.70–0.80 ppm is shown. Metabolite extraction and spectral measurement were

645 performed as described in Materials and Methods. The lyophilized leaf powder was
646 prepared from a local turnip variety *Kaida-kabu* harvested in Kiso district in 2015.

Table 1. Water-soluble compounds of *sunki* annotated by NMR analysis

#	Compound	#	Compound	#	Compound
	Aldoses, alditols		**Amino acids**		**Organic acids**
1	Glucose	11	Alanine	30	Lactic acid
2	Fructose	12	Asparagine	31	Acetic acid
3	Sucrose	13	Aspartic acid	32	Succinic acid
4	Glycerol	14	Glutamine	33	Formic acid
5	Mannitol	15	Glutamic acid	34	Propionic acid
6	*myo*-Inositol	16	Pyroglutamic acid	35	Phenyllactic acid
7	Erythritol	17	Glycine	36	4-Hydroxyphenyllactic acid
		18	Histidine	37	2-Hydroxyisovaleric acid
	Alcohols	19	Isoleucine	38	2-Hydroxyisocapric acid
8	Methanol	20	Leucine	39	Suberic acid/Pimelic acid*
9	Ethanol	21	Methionine		
10	2,3-Butanediol	22	Ornithine		**Amines**
		23	Phenylalanine	40	Ethanolamine
	Others	24	Proline	41	α-Aminobutyric acid
49	Acetoin	25	Serine	42	γ-Aminobutyric acid
50	Ascorbic acid	26	Threonine	43	Putrescine
51	Choline	27	Tryptophane	44	Cadaverine
52	Adenine	28	Tyrosine	45	Tyramine
53	Uracil	29	Valine	46	Tryptamine
54	Methiin			47	Histamine
				48	Phenethylamine

The compound numbers are corresponding to numerical labels for NMR spectra shown in Fig. 1 and Fig. S2.

*Not discriminated in this study.

Table 2. Peaks of volatile compounds of *sunki* annotated by SPME-GC/MS analysis

	Result of MS similarity search			
Peak #	Sample	RT (min)	SI*	Annotation
Esters				
1	A1	5.452	92	Ethyl acetate
7	A1	7.636	97	1-Propyl acetate
11	A4	9.830	97	Methyl thiolacetate
13	C1	10.637	97	Butyl acetate
16	A1	12.192	97	Isopentyl acetate
18	B3	14.070	96	Ethyl 4-pentenoate
22	A1	15.519	98	4-Penten-1-ol acetate
23	A1	15.629	97	2-Penten-1-ol acetate
24	B3	15.791	92	Methyl 5-hexenoate
26	A1	16.555	93	3,3-Dimethylallyl acetate
31	A1	18.506	96	3-Hexenyl acetate
33	G1	19.201	98	Ethyl lactate
51	A3	26.718	95	Methyl benzoate
54	A1	29.900	95	Geranyl acetate
55	A1	31.226	98	Phenethyl acetate
Ketones				
2	B1	5.674	98	2-Butanone
8	C3	7.660	95	2,3-Butanedione
19	E2	14.238	98	2-Heptanone
20	D3	14.508	96	2-Methylcyclopentanone
29	B1	17.478	97	Acetoin
38	E2	20.594	96	2-Nonanone
49	A3	26.043	97	Isophorone
Aldehydes				
3	F2	5.971	96	2-Methylbutanal
4	F2	6.076	96	3-Methylbutanal
46	A4	24.165	97	Benzaldehyde
52	F2	27.315	98	4-Methylbenzaldehyde
Alcohols				
5	E1	6.395	98	Isopropyl alcohol

6	A1	6.549	97	Ethanol
10	A1	9.164	98	2-Butanol
14	G1	11.640	96	Isobutyl alcohol
15	A1	12.104	98	3-Pentanol
17	A1	13.663	98	1-Penten-3-ol
21	G1	15.168	95	Isopentyl alcohol
25	B1	16.496	92	1-Pentanol
30	A1	18.024	95	4-Penten-1-ol
32	A1	18.593	93	2-Heptanol
37	A1	20.490	95	3-Hexen-1-ol
39	F2	21.113	97	2-Hexen-1-ol
41	C4	22.233	97	1-Octen-3-ol
58	A1	33.662	91	Phenethyl alcohol
Nitriles				
9	A1	8.333	98	Methyl isocyanide
28	A1	17.049	0.85**	4-Cyano-1-butene
34	A1	19.351	96	5-Cyano-1-pentene
35	A1	19.649	94	5-Methylhexanenitrile
43	A1	22.750	94	6-Cyano-1-hexene
44	B5	22.923	94	Octanenitrile
57	H1	33.652	80	5-(Methylsulfanyl)pentanenitrile
61	A1	35.618	97	Benzenepropanenitrile
Sulfides				
12	E2	10.558	95	Dimethyl disulfide
36	A4	20.330	96	Dimethyl trisulfide
ITCs				
27	B5	16.810	89	Butyl ITC
42	A1	22.406	86	3-Butenyl ITC
48	A1	24.641	0.86**	4-Pentenyl ITC
Acids				
40	A1	21.783	93	Acetic acid
45	B3	23.392	94	Formic acid
47	B1	24.328	98	Propanoic acid
50	B1	26.600	96	Butanoic acid

53	A1	27.665	97	2-Methylbutanoic acid
56	B3	31.634	98	Hexanoic acid
59	B3	33.855	92	2-Ethylhexanoic acid
60	B3	33.863	94	Heptanoic acid
62	A1	35.811	97	Octanoic acid

The peak numbers are corresponding to numerical labels for SPME-GC/MS chromatogram shown in Fig. 2.

*Similarity index calculated by the Shimadzu GCMS solution software.

**Similarity score calculated by MassBank spectrum search.

Table S1. List of *sunki* samples used in this study.

Factory	Sample ID	Year of production	pH	Na^+ (mg/L)
A	A1	2015	4.54	69
	A2	2015	4.05	73
	A3	2016	3.65	52
	A4	2016	3.63	51
B	B1	2015	4.18	47
	B2	2015	4.24	42
	B3	2016	3.72	46
	B4	2016	3.61	52
	B5	2016	3.65	51
C	C1	2015	4.06	29
	C2	2015	3.93	40
	C3	2015	3.80	41
	C4	2016	3.88	41
	C5	2016	3.87	41
D	D1	2015	3.73	30
	D2	2016	3.88	34
	D3	2016	3.92	45
E	E1	2015	4.06	27
	E2	2016	4.21	41
	E3	2016	4.11	42
F	F1	2015	3.93	96
	F2	2016	3.93	70
	F3	2016	3.91	71
G	G1	2015	3.76	22
	G2	2016	3.52	23
	G3	2016	3.58	24
	G4	2016	3.54	24
H	H1	2015	4.15	70

Fig.1

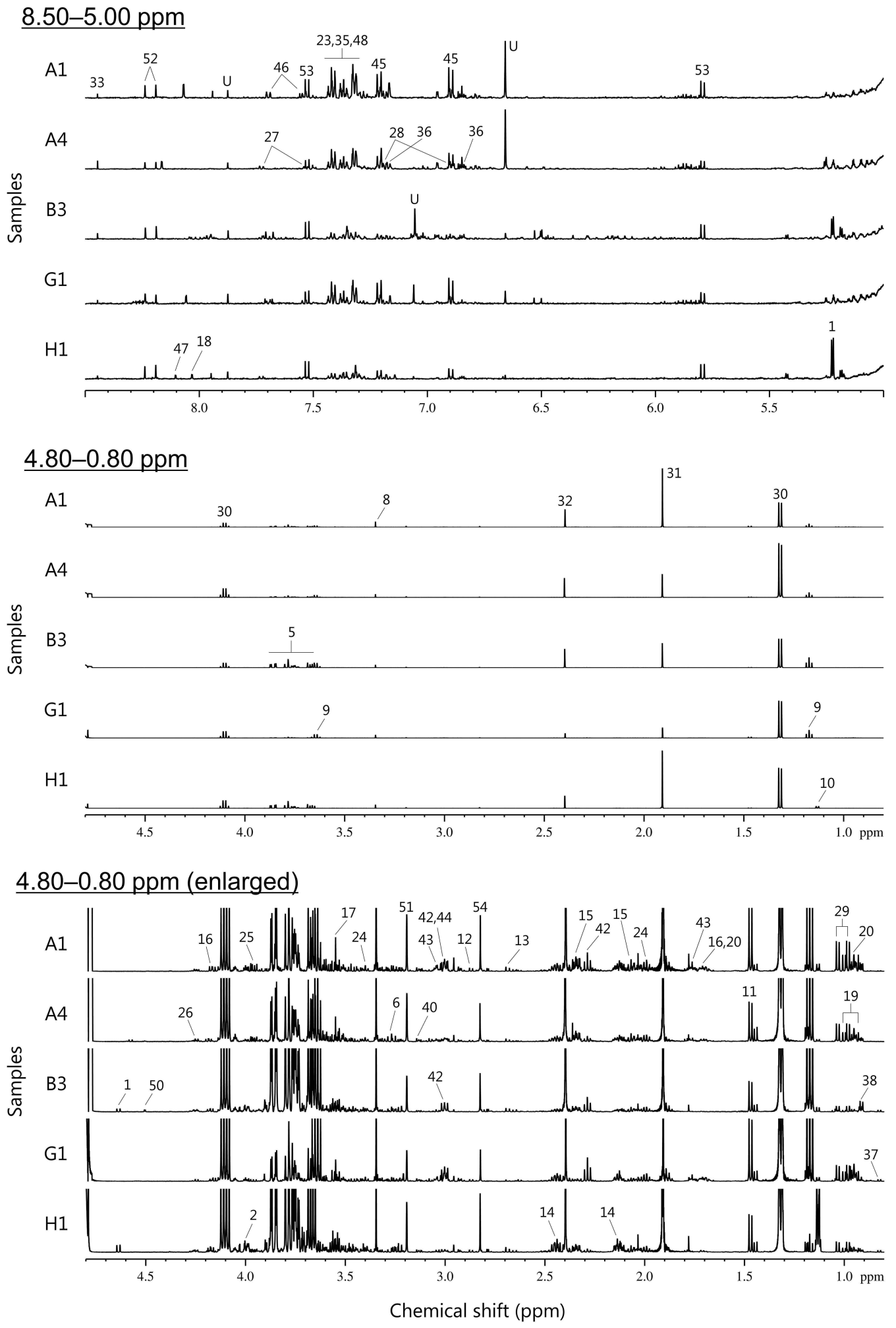

8.50–5.00 ppm
A1
A4
B3
G1
H1
Samples
33
52
U
46
53
23,35,48
45
45
U
53
27
28
36
36
U
47
18
1
8.0
7.5
7.0
6.5
6.0
5.5
4.80–0.80 ppm
30
8
32
31
30
5
9
9
10
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0 ppm
4.80–0.80 ppm (enlarged)
16
25
17
24
51
42,44
43
12
54
13
15
42
15
24
43
16,20
29
20
26
6
40
11
19
1
50
42
38
2
14
14
37
Chemical shift (ppm)

Fig. 2

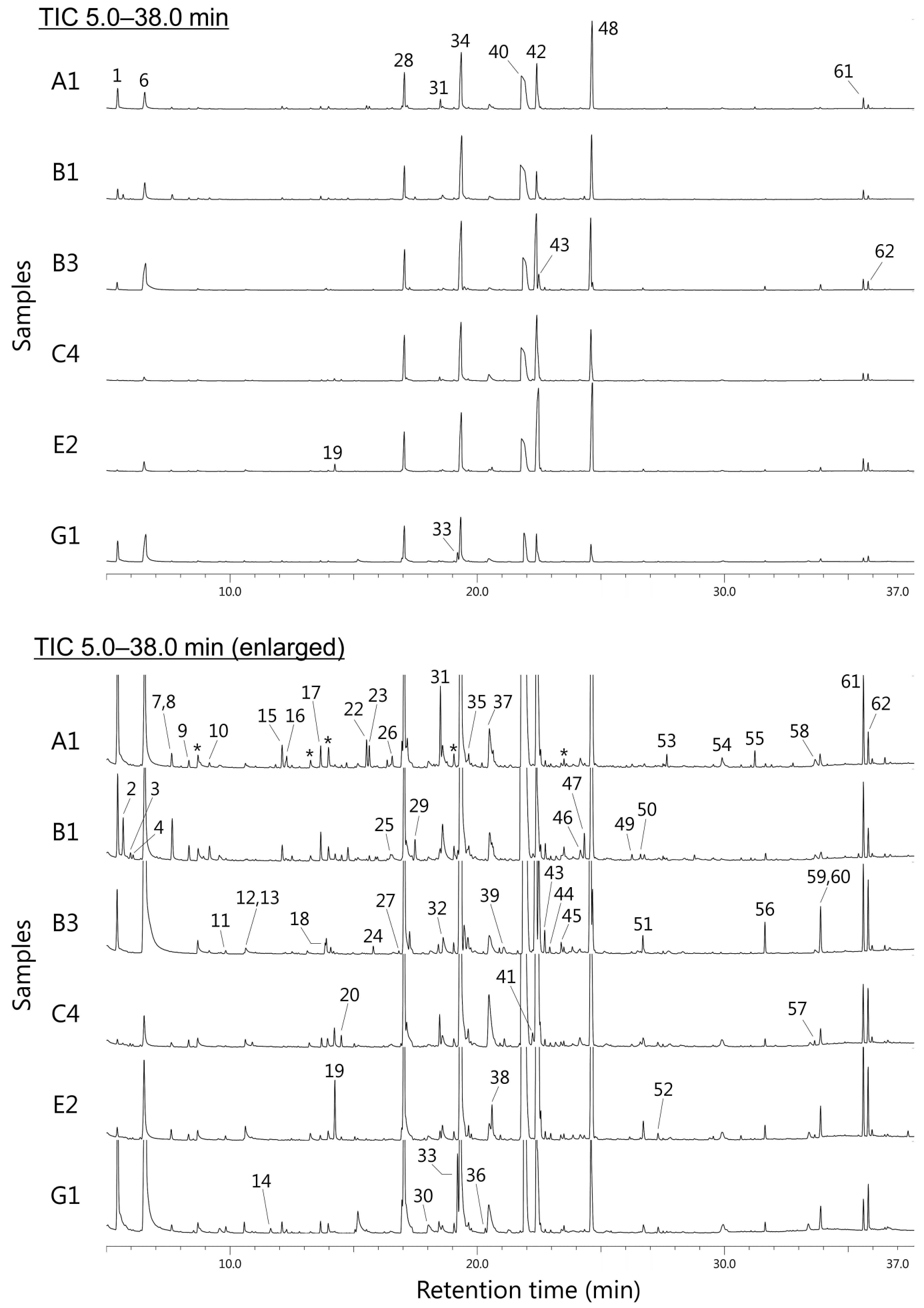

TIC 5.0–38.0 min
A1
B1
B3
C4
E2
G1
Samples
1
6
28
31
34
40
42
48
61
43
62
19
33
10.0
20.0
30.0
37.0
TIC 5.0–38.0 min (enlarged)
7,8
9
10
15
16
17
22
23
26
35
37
53
54
55
58
2
3
4
25
29
46
47
49
50
12,13
11
18
24
27
32
39
43
44
45
51
56
59,60
20
41
57
38
52
14
30
36
Retention time (min)

Fig. 3

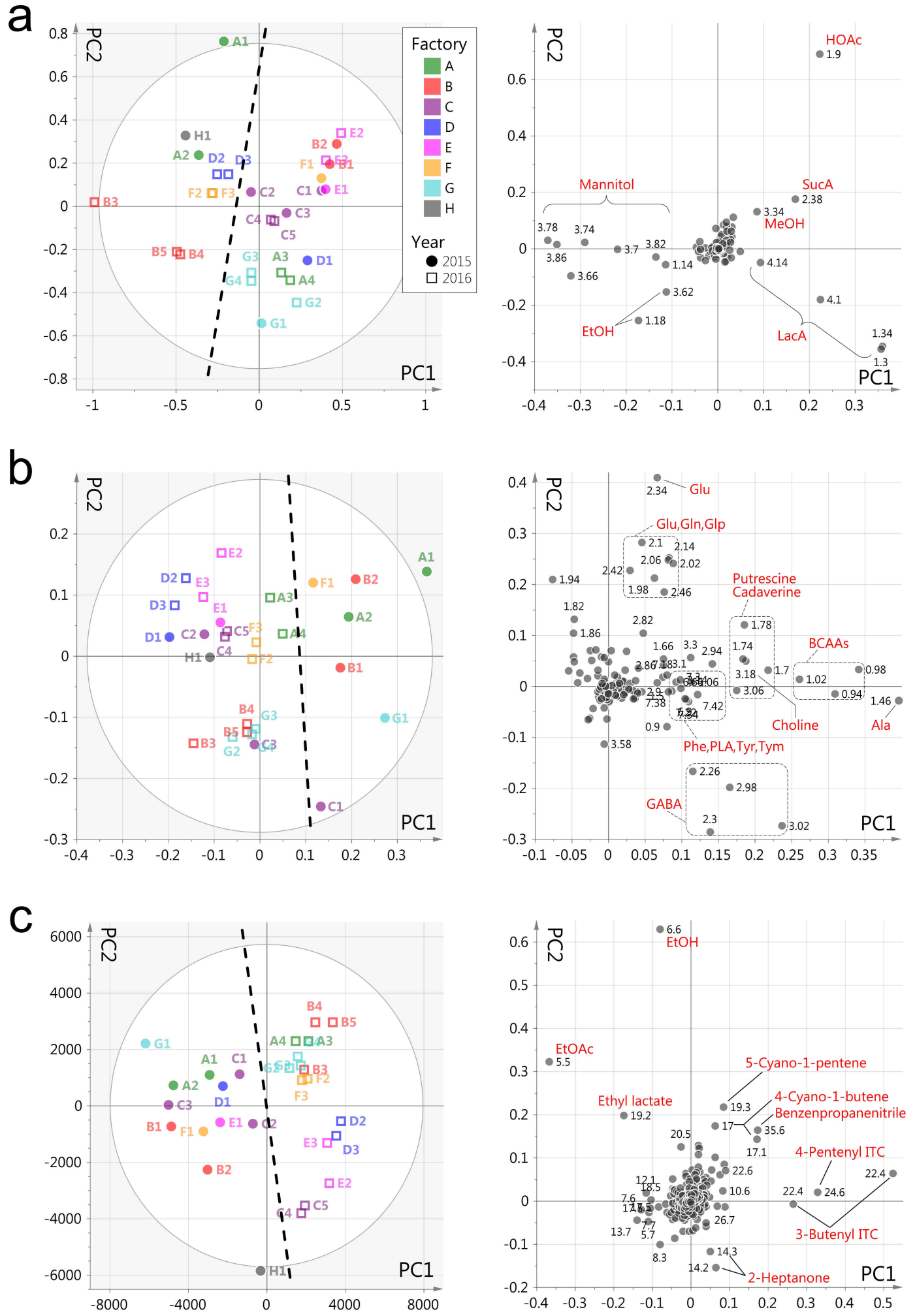

a
PC2
PC1
Factory
A
B
C
D
E
F
G
H
Year
2015
2016
A1
H1
A2
D2
D3
B2
E2
F1
F2
F3
B1
C2
C1
E1
B3
C4
C3
C5
B5
B4
G3
A3
D1
G4
A4
G2
G1
HOAc
1.9
Mannitol
SucA
2.38
3.34
MeOH
3.78
3.74
3.7
3.82
3.86
1.14
4.14
3.66
3.62
4.1
EtOH
1.18
LacA
1.34
1.3
b
PC2
PC1
E2
A1
D2
E3
F1
B2
D3
A3
E1
A2
C5
F3
A4
D1
C2
C4
H1
F2
B1
B4
G3
G1
B5
B3
G2
C3
C1
Glu
2.34
Glu,Gln,Glp
2.1
2.14
2.06
2.02
2.42
1.94
1.98
2.46
Putrescine
Cadaverine
1.82
2.82
1.78
1.86
1.66
3.3
2.94
1.74
BCAAs
3.18
1.7
1.02
0.98
3.06
0.94
1.46
Choline
Ala
7.38
7.42
0.9
3.58
Phe,PLA,Tyr,Tym
2.26
2.98
GABA
2.3
3.02
c
PC2
PC1
B4
B5
G1
A4
A3
A1
C1
G2
B3
F2
A2
F3
C3
D1
B1
F1
E1
C2
D2
E3
D3
B2
E2
C5
C4
H1
6.6
EtOH
EtOAc
5.5
5-Cyano-1-pentene
Ethyl lactate
19.2
19.3
4-Cyano-1-butene
Benzenpropanenitrile
17
35.6
20.5
17.1
4-Pentenyl ITC
22.6
12.1
18.5
10.6
22.4
24.6
22.4
7.6
26.7
13.7
5.7
3-Butenyl ITC
8.3
14.3
14.2
2-Heptanone

Fig. 4

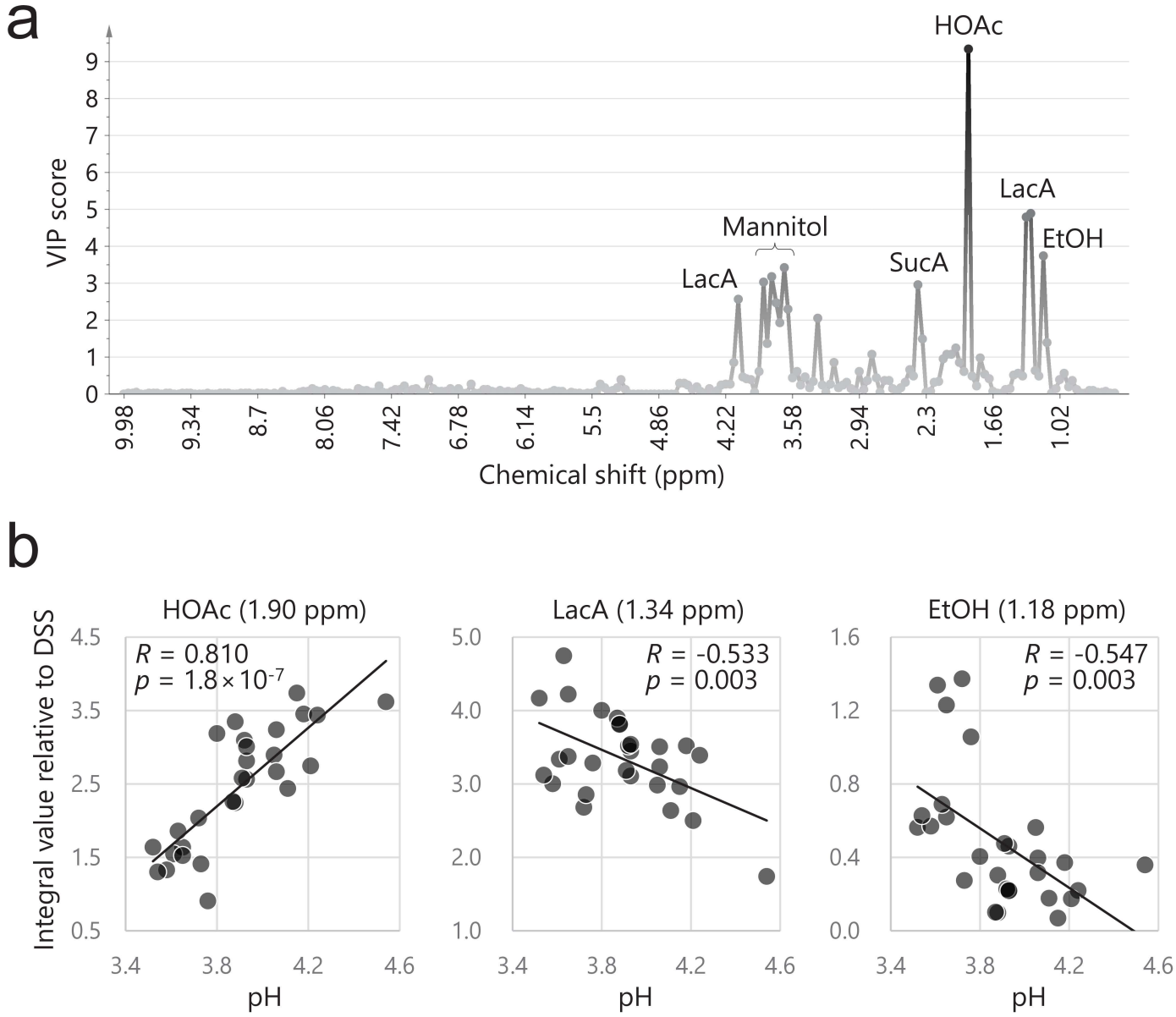

a
VIP score
Chemical shift (ppm)
HOAc
LacA
Mannitol
LacA
SucA
EtOH
b
HOAc (1.90 ppm)
R = 0.810
p = 1.8×10-7
LacA (1.34 ppm)
R = -0.533
p = 0.003
EtOH (1.18 ppm)
R = -0.547
p = 0.003
Integral value relative to DSS
pH

Fig. S1

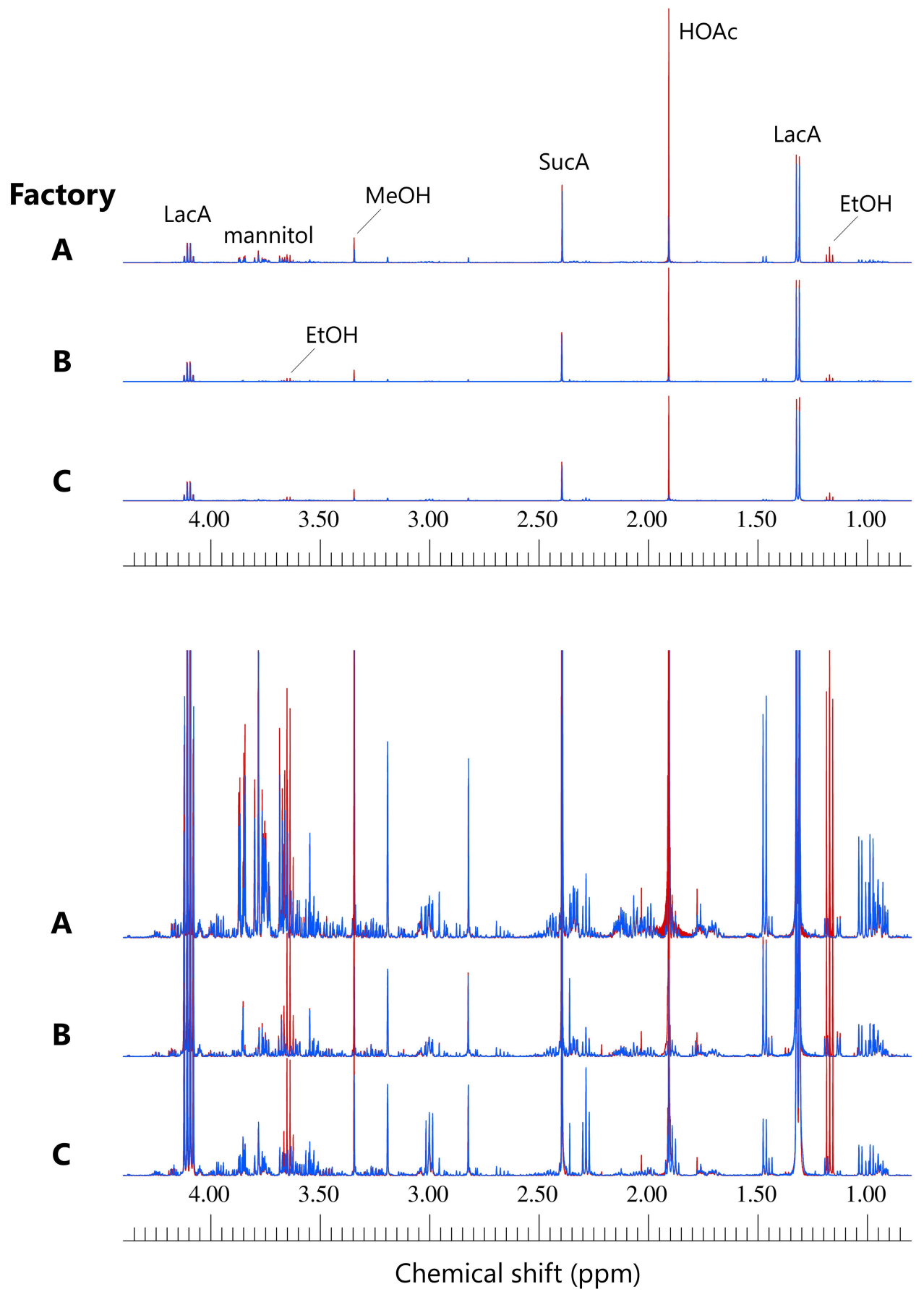

Factory
A
B
C
LacA
mannitol
MeOH
SucA
HOAc
LacA
EtOH
EtOH
4.00
3.50
3.00
2.50
2.00
1.50
1.00
A
B
C
4.00
3.50
3.00
2.50
2.00
1.50
1.00
Chemical shift (ppm)

Fig. S 2

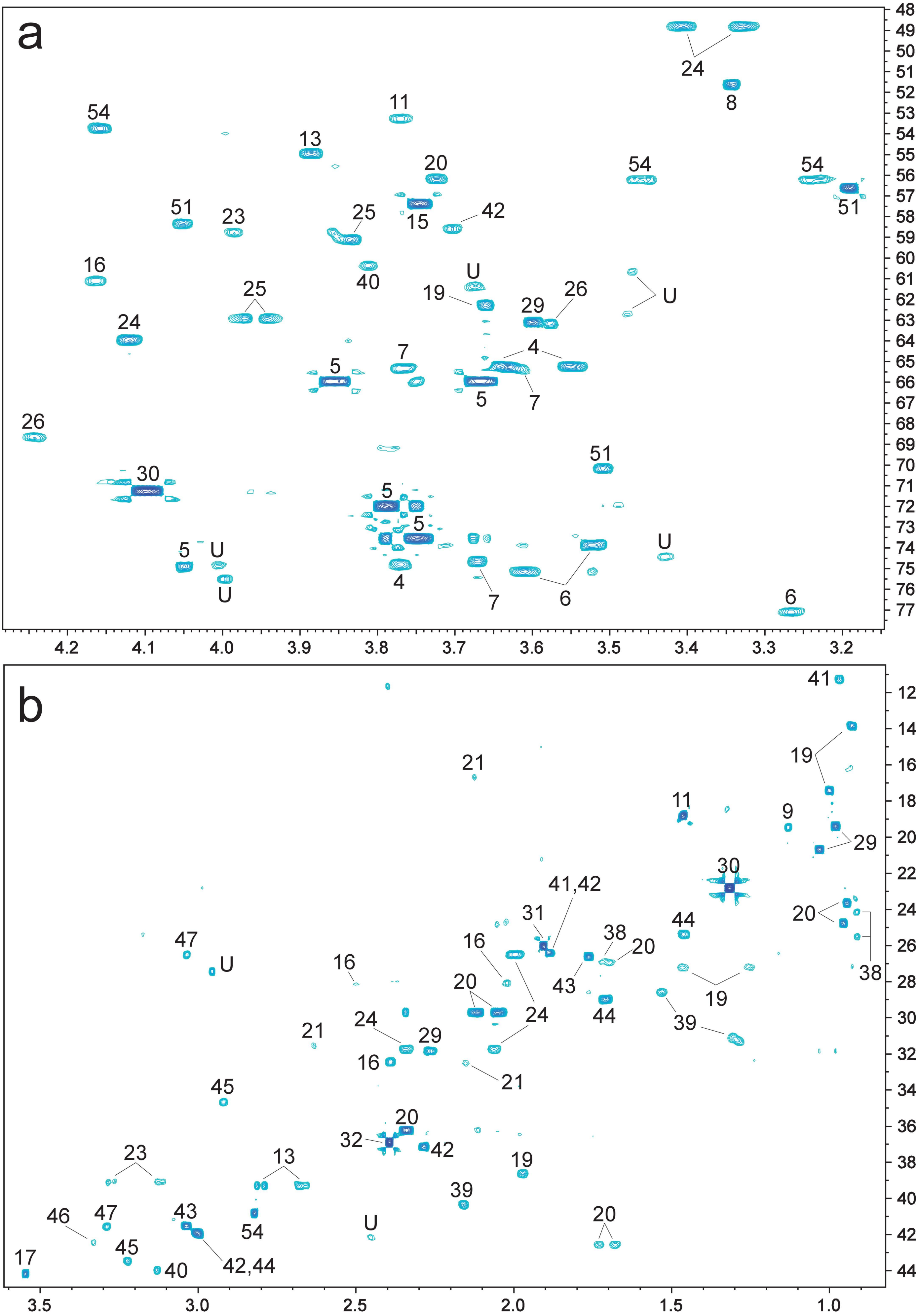
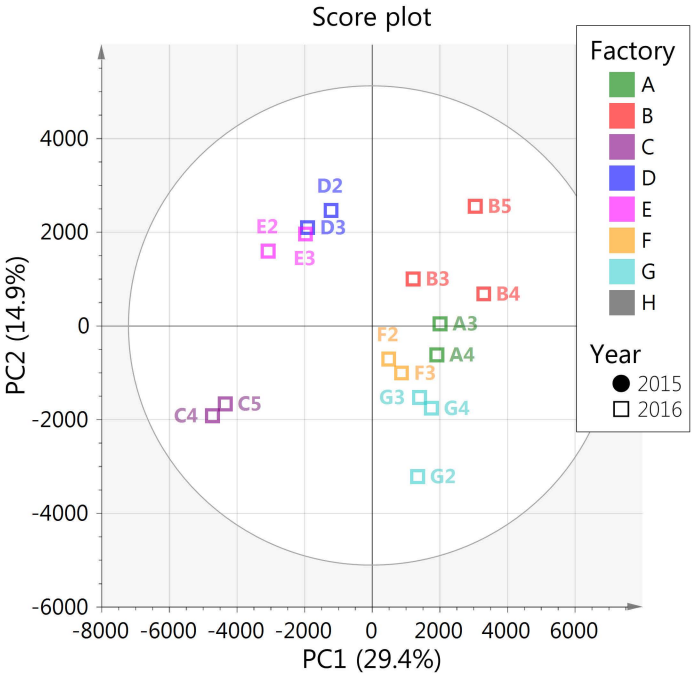
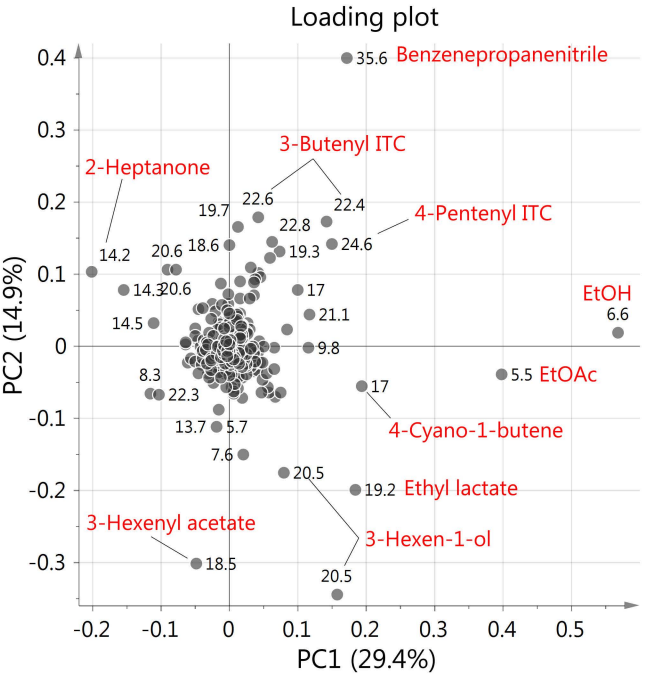

a
24
8
54
11
13
20
54
54
51
23
25
15
42
51
16
40
U
19
25
29
26
U
24
4
7
5
5
7
26
51
30
5
5
5
U
U
4
7
6
U
6
48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77
4.2 4.1 4.0 3.9 3.8 3.7 3.6 3.5 3.4 3.3 3.2
b
41
19
21
11
9
29
30
41,42
31
20
44
47
16
U
16
38
20
43
19
38
20
24
44
39
24
29
21
16
21
45
20
32
42
19
23
13
39
20
46
47
43
54
U
45
42,44
17
40
12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44
3.5 3.0 2.5 2.0 1.5 1.0

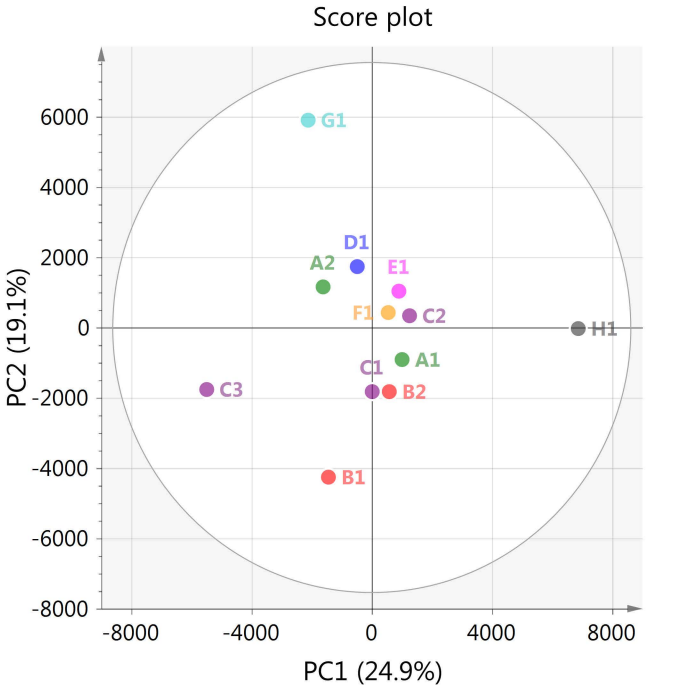
Fig. S3

GC/MS-based PCA for the samples in 2016





GC/MS-based PCA for the samples in 2015



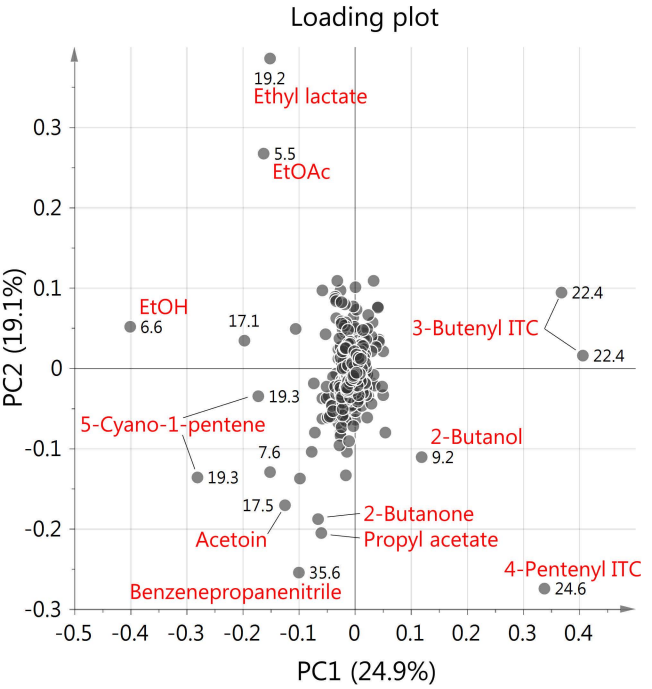
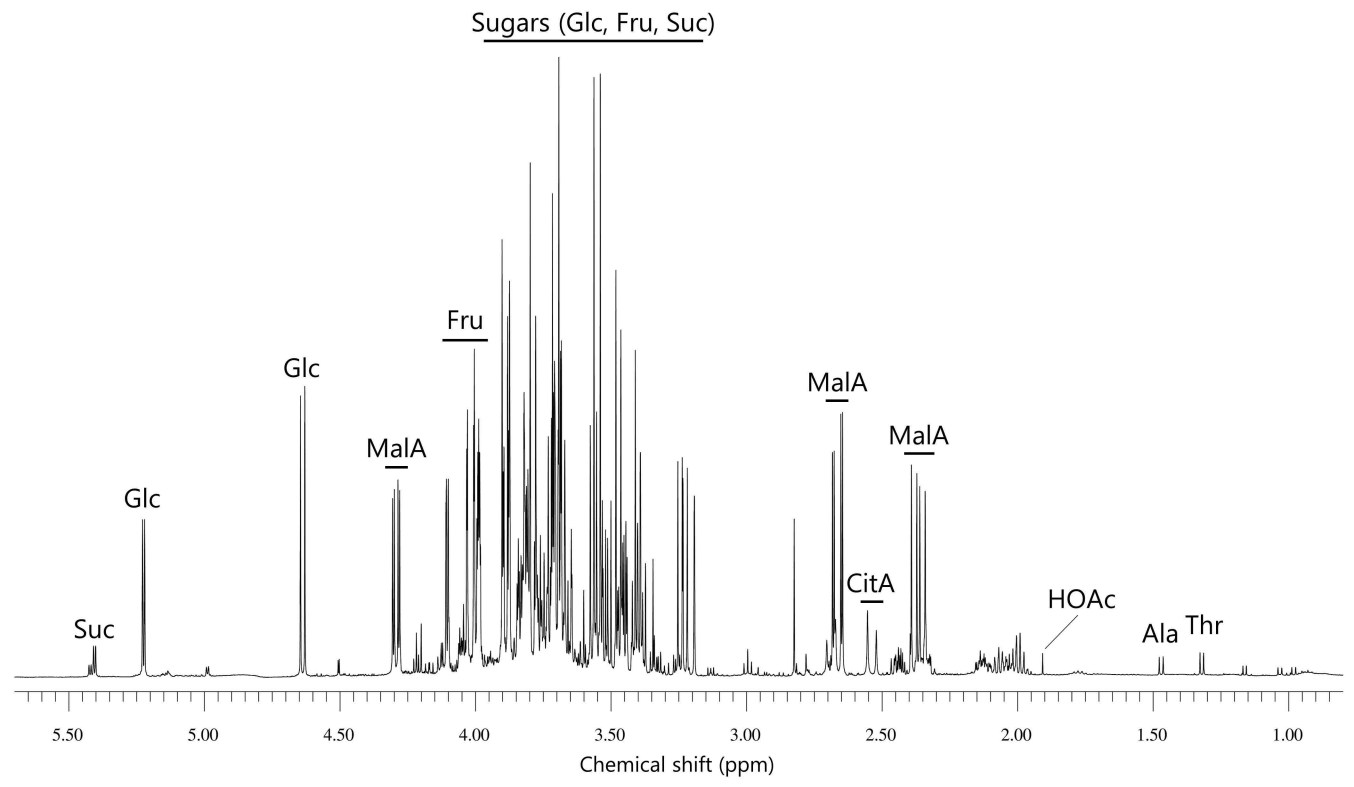


Fig. S4


Sugars (Glc, Fru, Suc)
Fru
Glc
MalA
Glc
MalA
MalA
CitA
HOAc
Suc
Ala
Thr
5.50
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
Chemical shift (ppm)