DR. OLUWAFEMI AYODEJI ADEBO (Orcid ID : 0000-0002-3757-5137)

DR. SAMSON OYEYINKA (Orcid ID : 0000-0002-5760-541X)

MS. JANET ADEYINKA ADEBIYI (Orcid ID : 0000-0003-0044-9126)

Article type : Review

Application of gas chromatography mass spectrometry (GC-MS)-based metabolomics for the study of fermented cereal and legume foods: A review

Oluwafemi Ayodeji Adebo[1*], Samson Adeoye Oyeyinka[2], Janet Adeyinka Adebiyi[1], Xi Feng[3], Jonathan D Wilkin[4], Yusuf Olamide Kewuyemi[5], Adrian Mark Abrahams[1], Fidele Tugizimana[6]

[1]Department of Biotechnology and Food Technology, Faculty of Science, University of Johannesburg, Doornfontein Campus, P.O. Box 17011, Gauteng, South Africa; [2]School of Agriculture and Food Technology, Alafua Campus, University of the South Pacific, Suva, Fiji; [3]Department of Nutrition, Food Science and Packaging, San Jose State University, One Washington Square, San Jose, CA 95192, USA; [4]Division of Engineering and Food Science, School of Applied Sciences, Abertay University, Dundee, United Kingdom; [5]School of Tourism and Hospitality, College of Business and Economics, University of Johannesburg, P. O. Box 524, Bunting Road Campus, Johannesburg, Gauteng, South Africa; [6]International R&D, Omnia Group, Ltd, P.O. Box 69888, Gauteng, South Africa

Corresponding author:* oadebo@uj.ac.za; +27115596261

Abstract

This article has been accepted for publication in International Journal of Food Science and Technology and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/ijfs.14794

A new era of cutting-edge technologies and advancements in analytical platforms and omics sciences is disruptively bringing a paradigm shift in fundamental and translational research. Metabolomics is one of the omics strategies that yields big data and has gained popularity in a wide spectrum of applications. Among various analytical platforms used in metabolomics, gas chromatography mass spectrometry (GC-MS) allows the measurement of thermally stable (volatiles and semi-volatiles) metabolites, with an advantage of spectral reproducibility. Cereal and legume-based fermented foods are part of the food culture in various countries throughout the world. Thus, this review provides an overview of recent applications of GC-MS-based metabolomics in the food fermentation field, specifically cereal and legume-based fermented foods. This emerging use of metabolomics in food fermentation studies illustrates the potentials of this omics science to elucidate metabolome landscapes of fermented foods. Such insights would advance our predictive understanding of fermentation processes and molecular descriptions of resultant food products; a necessary step for improvements and sustainability in food industry. Furthermore, the review echoes the current need of collaborative efforts in the scientific community (in this field) to harness and maximise the potentials of metabolomics in food fermentation studies.

Keywords: Fermentation, fermented foods, metabolomics, metabolites, GC-MS metabolomics, multivariate data analysis

1 Introduction

In the field of life sciences, one of the notable technological advancements is chromatography coupled with mass spectrometry. Over the last decade, this hyphenated analytical platform has been developed to increase reliability and sensitivity (Lesur *et al.*, 2016; Groger *et al.*, 2020). Chromatography coupled with mass spectrometry has become a central and widely used analytical system in metabolomics. The latter is a multidisciplinary science that aims to analyse the entire complement of small molecular weight molecules (≤1500 Da) within a biological or chemical matrix of interest. The application of metabolomics spans a wide range of fields including medicine, biological and life sciences, nutrition, agriculture, and more recently in food science and technology research (Tugizimana *et al.*, 2013; Adebo *et al*., 2017a; Adamski, 2020; Feng *et al*., 2020).

Particularly in food science research, the term foodomics (a fusion of two words, food and 'omics') was created in 2009 and refers to the study of food and nutrition through the application

of omics technologies (Cifuentes, 2009). This discipline encompasses a number of omics routes including proteomics, transcriptomics, and metabolomics used to unravel basic molecular food mechanism in relation to health (León *et al*., 2018; Li *et al*., 2020). Advancement in this regard is expanding existing knowledge in different food sectors including food safety and microbiology, food processing, food microbiology, food traceability and food authenticity, food contamination and fraud, and within food functionality. In this context, metabolomics has been applied as a reliable analytical approach to elucidate the global biochemical changes and biotransformation of metabolites in food processes, such as fermentation (Adebo *et al*., 2017a; Singh *et al*., 2017; Park & Kim, 2019; Wang *et al*., 2020). The latter is an age-long food processing technique involving the transformation of substrates through microbial activities (Adebo, 2020). Such a process leads to several modifications that could alter metabolite levels and/or synthesize new ones (Kohajdova & Karovicova, 2007; Adebo & Medina-Meza, 2020; Kewuyemi *et al*., 2020a).

Cereals and legumes are major staple crops and primary source of nutrients to millions of people all over the globe (Patil *et al*., 2016; Oghbaei *et al*., 2016). These food groups are frequently fermented into alcoholic and non-alcoholic beverages, gruels, porridges, etc. (Blandino *et al*., 2003; Adebo *et al.,* 2017b; Adebo, 2020). Depending on the substrate, typical metabolites could include but are not limited to flavor/aroma related constituents (volatiles), products of proteolytic actions (amino acids and peptides), alcohol-related compounds, starch and carbohydrate fractions, and secondary metabolites (including phytochemicals) (Verbeke *et al*., 2015; Adebo *et al*., 2017a; Adebo *et al*., 2019; Raghuvanshi *et al*., 2019). Understanding such a complex and multidimensional metabolic space with diverse concentrations, chemical structures, affinity, and polarities, can be somewhat challenging using classical and conventional techniques. Hence, metabolomics – involving the global qualitative and quantitative profiling of metabolites in a biological matrix – can provide a comprehensive characterisation of the products of food processes. This positions metabolomics as a desirable approach for providing better insight as well as understanding the multifunctionality and complexities of cereal and legume fermented foods (CLFFs).

Capillary electrophoresis-mass spectrometry (CE-MS), gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS), and nuclear magnetic resonance (NMR) are the most frequently used analytical platforms in metabolomics studies (Adebo *et al*., 2017a; Adamski, 2020; Ten-Doménech *et al*., 2020). Other analytical platforms used in

metabolomics include Fourier Transform Infrared (FTIR) and Raman spectroscopy. GC-MS platform, the focus of this review, has been widely used in metabolomic studies. The analytical aspects of current GC-MS technologies worth noting include high resolution (in both GC and MS technologies), advancements and improvements in ionisation processes, reproducible fragmentation pattern (which is an advantage for compound identification) as well as fewer matrix effects. Thus, in this review, an overview on the use of GC-MS-based metabolomics for the study of CLFFs is appraised and discussed. Focus was mainly on recent studies CLFF-GC-MS-based metabolomics studies between 2010 – 2020.

2 Fundamentals of metabolomics

2.1 Metabolomics approaches and study design step for fermented foods

Metabolomics, like other 'omics' sciences, employs either targeted or untargeted approaches. The former is often a hypothesis-driven approach, focusing on a pre-defined class or specific group of metabolites and often with an absolute quantitative description of measured metabolites. Untargeted metabolomics on the other hand, is a data-driven approach, and aiming at a wide coverage of the metabolome under consideration, generating metabolic patterns with a relative quantitative description of the measured metabolite (Cevallos-Cevallos *et al.*, 2009; Mozzi *et al*., 2013; Tugizimana *et al*., 2013; Adebo *et al*., 2017a). The intended biological question of the study determines the study design and choice of approach to follow. Theoretically, an untargeted approach leads to the generation of hypotheses, which can then be investigated further with a targeted study, focusing on selected metabolites or class of specific metabolites.

Experimentally, irrespective of the analytical platform used, a metabolomics study follows a general multi-step workflow (Figure 1). Furthermore, being at the interface between chemistry, biology, data science, chemometrics, and computer science, metabolomics is a multidisciplinary field, which makes metabolomics study a team effort (Tugizimana *et al*., 2013). With a biological or research question in mind, the study and experimental design is the first step of metabolomics workflow through which key aspects of the downstream steps are critically assessed and designed – defining the (biological) system under consideration (e.g. CLFFs), metabolomics approach to use (untargeted *vs*. targeted), statistical considerations (e.g. sample size, number of samples per groups, batches and replicates), extraction method(s) to apply, an analytical platform to use, planning of the data mining methodologies, and timeframe of the study (Tugizimana *et al*., 2013). This experimental design step is critical and essential to ensure the success of the study whilst

minimising errors, ensuring a necessary scientific rigour and meeting minimum requirements (as defined by the Metabolomics Standard Initiative) (MSI, 2005). In carrying out the downstream experimental steps, it is pertinent to minimise sources of unwanted variation using standard experimental operating procedures across biological replicates and batches, at both pre- and during analytical steps (Tugizimana, *et al*., 2013; Adamski, 2020). Detailed descriptions of the metabolomics multistep workflow have been extensively covered and presented in several literatures, some of which are cited herein (Villas-Bôas *et al*., 2005; Dettmer *et al*., 2007; Álvarez-Sánchez *et al*., 2010; Gu *et al*., 2011; Dunn *et al*., 2012; Dunn & Hankemeier, 2013; Tugizimana *et al*., 2013; Fiehn, 2017). However, it may suffice here to provide an overview of some of these metabolomics workflow steps, highlighting key considerations and critical aspects for a successful metabolomics work.

2.1.1 A pre-analytical step: extraction of metabolites and sample preparation

Once the study system has been confirmed and all aspects mentioned as previously outlined above, metabolites to be measured must be extracted and prepared for the analytical step. Since CLFFs could come in various forms of solids, semi-solids or liquid forms, aspects relating to experimental sample preparation include pulverising, homogenisation, lyophilisation as well as drying (e.g. freeze drying, liquid nitrogen) after which metabolites (or analytes of interest) are extracted (Adebo *et al*., 2017a; Hyeon *et al*., 2020).

Extraction is essentially needed to release metabolites from the (biological) matrix and may require optimisation during targeted analysis (to reduce interferences of unwanted chemical species) and in untargeted analysis (to improve metabolite coverage metabolites) (Roberts *et al*., 2012; Gbashi *et al.*, 2017; Wang *et al*., 2019). Several extraction procedures such as direct solvent extraction, liquid-liquid continuous extraction, QuEChERS (quick, easy, cheap, effective, rugged, and safe) method, simultaneous steam distillation and extraction, solid-phase microextraction (SPME), headspace, solid-phase extraction (SPE), solvent-assisted flavor evaporation (SAFE) among others, are employed for the release of metabolites from CLFF matrices (Jo *et al*., 2011; Adebo *et al*., 2019; Mu *et al*., 2019; Wang *et al*., 2020).

While the QuEChERS method illustrates a single-step acetonitrile extraction and dispersive SPE, the direct solvent extraction involves blending analyte with an organic solvent (acetonitrile, acetone, ethyl acetate, chloroform, ethanol or methanol) (Lee *et al*., 2014; Lorenzo & Pico, 2017). Both solvent extraction processes often require further derivatisation step to increase analyte

volatility, for analyses on a GC-MS system (Lee *et al*., 2016; Yin *et al*., 2017; Hyeon *et al*., 2020). Through common chemical derivatisation reactions (acylation, alkylation, or silylation) applied in GC analysis, the derivatised compounds become less polar moieties and could thus be eluted from a GC column with enhanced separation and sensitivity (Dettmer *et al*., 2007; Cooray & Chen, 2018). For SPME, divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) is the commonly coated fiber used in CLFFs, for volatile compounds adsorption (Park & Kim, 2019; Saa *et al*., 2019). This SPME-based method is solvent-free and offers high sensitivity (Balestra *et al*., 2015). As such, SPME is well-established with GC-MS (SPME-GC-MS) and provides good reproducibility, high-throughput analyses, and better metabolite detection for volatiles (Iijima, 2014; Lee *et al*., 2019). At relatively low temperature (40±20 ^{o}C), SAFE involves the vaporisation of volatiles from non-volatile analyte under ultra-high vacuum (Jo *et al*., 2011). However, not all metabolites in CLFFs can possibly be acquired and detected using a single extraction process and analytical platform. Only a multi-extraction and multi-platform approach can provide an ideal comprehensive coverage and holistic characterisation of metabolites present in CLFFs.

2.1.2 Analytical step: data acquisition, sample analysis

Data acquisition in the form of separation and subsequent detection of metabolites follows the metabolite extraction and sample preparation step. In this case of GC-MS-based metabolomics approach, while the separation of sample components is done using GC, the detection of these metabolites is chiefly accomplished through an MS system. Quality control and assurance (QC and QA) for sample analysis step are critical considerations to ensure the quality of the generated data: *i.e.* to minimise (or eliminate) analytical bias, background noise, unwanted (analytical) variation, artefacts. The QA and QC step also ensures stability, reproducibility and repeatability. More details on practical guidelines in this regard are available in the literature (Bouhifd *et al*., 2015; Broadhurst *et al*., 2018; Barnes, 2020). GC-MS systems (in which electron impact / EI ionization method is used) generates reproducible ion fragmentation patterns, which made it possible to build spectral libraries for metabolite identification such as the Feihn metabolomics library, combined chemical dictionary (Chapman and Hall/CRC), Golm metabolome database, Wiley library (W8N05ST; Agilent Technologies Inc.), NIST/EPA/NIH National Inst. of Standards and Technology (NIST) library, Mass Bank as well as Palisade (Dettmer *et al*., 2007; Fiehn, 2017) making it an integral tool for metabolite identification, providing analyte-specific detection, high-chromatographic metabolite resolution and quantification, as well as the capability to identify unknowns (Dettmer *et al*., 2007; Fiehn, 2007).

The technological advancements in GC-MS instrumentations has led to the emergence of a comprehensive two-dimensional (2D) GC (GC × GC) hyphenated with modern high resolution mass spectrometry, having high sensitivity in full-spectrum-acquisition mode, and fast scan speeds (Cajka, 2013; Adebo *et al*., 2019). Furthermore, current MS systems coupled to GC platforms are also available as hybrid systems, combining different mass analysers, such as triple quadrupole (QqQ), Q-Orbitrap, TOF-MS, and Q/TOF-MS with high mass accuracy measurement and multi-level fragmentation capabilities (Adahchour *et al*., 2005; Lorenzo & Pico, 2017). The application of these modern mass spectrometric techniques for target and untargeted analysis of volatile/semi-volatile analytes in CLFFs, combined with the use of relevant spectral libraries (e.g. those mentioned above), has provided a better and comprehensive characterization of metabolites present in CLFF samples.

2.1.3 Post-analytical step: handling and mining metabolomics data

Following data acquisition on a GC-MS system, the generated data are mined through a multistep pipeline that comprises data extraction, pre-processing, pre-treatment and application of chemometrics and statistical methods. Data mining, a crucial and essential step in metabolomic workflows, can be carried out in two different approaches: (i) the chemometrics approach, in which the metabolites are not firstly identified (or annotated), but their spectral patterns are statistically evaluated to extract relevant spectral features that related to key questions of the study; and (ii) targeted profiling approach, in which most of the metabolites are firstly annotated (or identified) and then various statistical methods are applied to extract information related to the study. The choice of the approach to follow would depend on the study design and availability of resources.

Data pre-processing methods includes noise filtering, peak detection and peak alignment. In addition, data pre-treatment or data correction comprises data normalisation, centering, scaling, batch effect correction and data integrity checking (Tugizimana *et al*., 2013). Both pre-processing and pre-treatment assist in data cleaning to emphasise only relevant information. These steps inevitably determine the quality and quantity of the information obtained and subsequently the knowledge acquired. For chemometrics and statistical modelling, multivariate methods (both unsupervised and supervised approaches) are applied. These include the classical principal component analysis (PCA) – an unsupervised method for dimensionality reduction and data exploration, and hierarchical clustering analysis (HCA) for sample clustering. Furthermore,

supervised multivariate methodologies such as partial least square (PLS) and its variants (e.g. partial least square discriminant analysis (PLS-DA), orthogonal partial least square discriminant analysis (OPLS-DA)) are applied to interrogate the data in supervised manner for sample classification and extracting significant metabolites that discriminate sample groups (Tugizimana *et al*., 2013).

There are currently several public and license-based software tools, resources and bioinformatics platforms available for this step of data mining and interpretation. Some of the tools commonly used (and applicable for CLFF-metabolomic data) include ChromaTOF, Mass Hunter, XCMS, MetaboAnalyst, KEGG and ChemStation.

2.2 Some advantages and limitations of GC-MS platform in metabolomics

According to Lu *et al*. (2018) and Fiehn (2017), due to the availability of standardised libraries and constant electron ionization used to accumulate mass spectra and chromatographic retention times of over 500, 000 compounds, GC-MS platform is regarded as the gold standard analytical system in metabolomics, particularly for small and thermally stable metabolites. In GC-MS-based metabolomics, due to less ionisation artefacts and reproducible ion fragmentation, peak picking and spectral similarities search (against spectral databases) can be achieved with high confidence (Qualley & Dudareva, 2009; Fiehn, 2017). Subsequently, there are computational tools available for GC-MS-derived spectral data, such as automated mass spectral deconvolution software (AMDIS), which may not be readily available for LC-MS platforms (Lee *et al*., 2012). Other advantages of GC-MS worth pointing out (though arguably) include minimal ion suppression and matrix effects, and a relatively ease of use, in terms of analyses time and operating costs ease of use (in terms of analyses time and operating costs) (Gowda & Djukovic 2014; Mastrangelo *et al*., 2015; Fjeldsted, 2016; Lorenzo & Pico 2017; Beale *et al*., 2018).

Furthermore, as highlighted in Section 2.1.1, not all metabolites in CLFFs can be directly measured by conventional GC system, hence thermally unstable compounds such as some primary metabolites commonly found in CLFFs such as sugars, organic acids and amino acids requires an additional derivatisation step. However, derivatisation has its challenges in that subsequent detection and measurement is based on using the derivative as a proxy for the target compound. Hence, ensuring the completion of the derivatisation reaction as well as using the right

concentration of derivatising agent is essential for optimal transformation of the analyte into the derivatised form for GC measurement (Beale *et al*., 2018).

Despite the advantages of GC-MS highlighted above and those that can be found in the literature cited herein, this analytical platform also has some inherent limitations. The latter include the mass range of ca. 50–600 Da, limiting the metabolome coverage, and a sample preparation step that can be laborious if derivatisation is included. For instance, mass range covered by LC-MS systems is wider compared to that of GC-MS platforms, and for NMR and FTIR (though less sensitive), there is less sample preparation (or no metabolite extraction) needed. NMR platforms though less sensitive compared to MS systems, are nondestructive and offer an unbiased assessment of a complex sample, allowing the simultaneous identification and quantification of diverse metabolites (Simler *et al*., 2016).

3 GC-MS based metabolomics of CLFFs

The process of food fermentation is an age-long practice, generally aims to convert edible substrate into improved products by the action of microorganisms; naturally/spontaneously induced or starter aided (Kewuyemi *et al*., 2020b). Particularly for CLFFs, the main microbes involved in this process are lactic acid bacteria, molds, and yeasts. These fermentative microbes synthesise diverse active-intracellular enzymes which stimulate the bioconversion of macromolecules (carbohydrates, proteins, and lipids) into beneficial metabolites such as volatile compounds (amino acid, fatty acid, peptides, phenols etc.) with enhanced substrate properties including better nutritional composition, sensory, and functional properties (Park & Kim, 2019; Adebo & Medina-Meza, 2020; Kewuyemi *et al.,* 2020a). Cereals and legumes are typical substrates for fermentation and present readily available functional and nutraceutical benefits (Adebo *et al.,* 2017b; Verni *et al.,* 2019; Adebo, 2020). Understanding the beneficial and complex transformation in their preparation has necessitated the growing application of GC-MS-based metabolomics for the investigation of CLFFs (Table 1). Various studies in this context are discussed in the ensuing sub-sections.

3.1 Cereal-based FFs

As exemplified in Table 1, metabolomics can indeed aid in understanding the fermentation process of foods. Recent studies have demonstrated the robustness of this 'omics' approach for CLFFs by investigating varying biological question to present detailed insight of fermentation process. More specifically, the untargeted route was well-established to describe the fermentation process of

cereal-based foods (alcoholic and non-alcoholic beverages, bread, dough, gruel, among others). Seo *et al.* (2016) described the metabolic changes in *Makgeolli* (a traditional fermented rice wine) during alcoholic fermentation and aging using GC-MS based untargeted metabolomics to determine the fermentative behaviours of yeast strains. The study reported that fermentation progressed rapidly during the early fermentation period and decreased levels of glucose and phosphoric acid was observed while other identified metabolites increased. The observed metabolite changes were attributed to the different fermentation behaviours induced by the cultured yeast strains. On the other hand, during the aging period (up to 70 days), metabolites present in the product barely changed. A similar metabolomics approach was adopted by Mu *et al.* (2019) to investigate the relationship between metabolites and fermentation time of black glutinous rice wine. Through liquid extraction, subsequent derivatisation [methoxy amination and addition of N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA)], analysis on gas chromatography time of flight mass spectrometry (GC-TOF-MS) system and applying multivariate data analysis (MVDA), the authors reported the presence of 28 significantly different metabolites (SDMs) and 50% of these metabolites (phenolic acid, organic acid, and sugar) where enriched at 60 h. Using pathway analysis, the authors also revealed that alanine, aspartate, and glutamate metabolism, starch and sucrose metabolism, and pentose phosphate pathway were most relevant to pre-fermentation, with 60 h identified as the potentially optimal time for pre-fermentation of black glutinous rice wine (Mu *et al*., 2019).

Another study on rice wine characterized the volatile organic compounds (VOCs) during Chinese rice wine aging (Wang *et al.,* 2020). The study identified 94 VOCs as aging markers for discrimination of short-aged (0-3 years) rice wines (alcohols, phenols and their derivatives, sulfides, small esters and acids) and long-aged (5-15 years) wines (aldehydes, aromatics, furans, ketones, most acids and esters). Song *et al.* (2020) demonstrated untargeted and targeted analyses to identify potential markers for the classification of volatile composition in strong aroma-type *baijiu* samples (distilled liquor) from different geographical origins. Their findings indicated 29 potential markers and 22 marker compounds were selected for distinguishing the liquor samples. The discrimination ability was closely correlated to the characteristic flavour compounds (acid, alcohols, esters, furans, sulfides, and pyrazine). The targeted approach also revealed the markers had great discrimination power to differentiate the *baijiu* samples and connected the geographical origin to the composition of *baijiu* samples. Yin *et al.* (2017) reported 62 intracellular metabolites of lager beer flavour compound synthesised on an industrial scale. The most dominant metabolite

group identified was amino acids and concentrations recovered decreased towards the end of fermentation, probably assimilated by the fermenting yeast for synthesis of aroma-active metabolites at different fermentation stages.

By employing GC-MS based metabolomics, Ferri *et al.* (2016) examined the effect of two wheat flours and different *Lactobacillus plantarum* strains on the flavouring and antioxidant characteristics of fermented sourdough. Their result for KAMUT khorasan wheat-fermented dough showed a high correlation between a group of volatiles (alcohols, carbonils, dodecanoic acid, and 1,3-hexadiene) and polyphenolic compounds. For durum wheat sourdough, a different pattern of volatile molecules (acids, alcohols, carbonils, and hydrocarbons) was highly correlated with epigallocatechin, epigallocatechin-gallate, flavonoids, and total polyphenols. In addition, a simultaneous increase of sensorial and health beneficial compounds of the dough was dependent on specific combination of *L. plantarum* strains and wheat flour type. Both the mature and immature grains of these two wheat types were fermented by a sourdough made of *L. spp.* and *Saccharomyces cerevisiae* to investigate the volatilome of sourdoughs (Saa *et al.,* 2019). Using SPME-GC-MS-based metabolomics, the study indicated that sourdough process generates different volatile compounds compared to industrial fermentation (reference sample). Specifically, the volatilome of sourdough KAMUT khorasan including short chain fatty acids was most promising.

The development of yeast-free wheat doughs with different content of glucose and NaCl and fermentation by comparatively using *Zymomonas mobilis* strains had been demonstrated (Nissen *et al.,* 2020a). These authors reported that *Z. mobilis* strains produced about 10 mg ethanol/g dough, with maximum dough volumes (640-680 mL) attained after 2 h leavening while the presence of NaCl in the dough reached comparable values after 6 h. They also recorded unique signatures of the strains for the production of nonanoic and undecanoic acids, 2-hexadecenal, (E), L(+)-tartaric acid diethyl ester, and 3-decen-5-one, 4-methyl, (E). Nissen *et al*. (2020b) evaluated the effect of addition level of hemp seed flour and sourdough fermentation on the production of VOCs in gluten-free bread. In comparison to standard breads, the result of the study showed an increased concentration of antimicrobial compounds, a typical flavouring profile, and a larger spectrum of bioactive VOCs. Also, an increased in fermentation metabolites, mainly, acetic acid, ethanol, 1,4-butanediol, and 2-butanone-3-hydroxy were also observed.

The metabolite profile of rice *koji* produced using *Aspergillus oryzae* and *Bacillus amyloliquefaciens* at different fermentation times was carried out by Lee *et al.* (2016). They reported differences in the *koji* samples with respect to their fermentative species and duration of fermentation. In particular, rice *koji* fermentation by *A. oryzae* was associated with carbohydrate metabolism intermediates, fatty acids, and serine-derived amino acids whilst rice *koji* induced by *B. amyloliquefaciens* was linked to the presence of aromatic and branched chain amino acids, lysophospholipids, and flavonoids. The differential volatile metabolite profiles of two WG-sorghum types (high tannin and low tannin) and derived fermented products (WG-*ting*) obtained via controlled and spontaneous fermentation was descriptively elucidated by Adebo *et al*. (2019). SDMs reported include cyclic compounds, esters, fatty acid derivatives, ketone, organic acids, pesticides, phenol and a sugar derivative. Thus, the study demonstrated that the inherent metabolic profile of raw sorghum led to differential metabolic changes in WG-*ting* and could subsequently have dietary and health implications.

An attempt to understand the role of fermentative microbes in volatile metabolite formation focused on diverse approaches to compare metabolic characteristics and determine microbe-specific metabolites in fermented rice by diverse lactic acid bacteria, molds, and yeasts (Park & Kim, 2019). The study findings revealed that metabolic changes in fermented rice via molds inoculation were relatively more activated compared to other microbes. The correlation analysis of the volatile compounds in fermented rice with specific microbes indicated that the branched-chain volatiles were closely associated to *Aspergillus oryzae* whereas acetic acid had strong relationship with *Lactobacillus plantarum*. Kum *et al.* (2015) investigated the volatile profiles of rice-*koji doenjang* inoculated with three types of *Aspergillus* species and fermentation was done over a range of durations. The early phase fermentation of the samples was reported to be dominated by carbonyls while at the latter stage of fermentation long-chain fatty acid esters were considerably enhanced. The formation of the fatty acid derivatives was suggested as distinctive flavour components of rice-*koji doenjang*. Fermentation as a value addition process was used to process brewer's spent grain (BSG) inoculated with food grade *Rhizopus oligosporus* (Cooray & Chen, 2018). The metabolite variations of the fermented BSG showed significant increased level of changes in amino acids, antioxidants, citric acid and vitamins, thus, such insight of metabolite changes could pave way for novel application.

3.2 Legume based FFs

Most studies have channeled considerable attention to the elucidation of metabolite changes in fermented soybean products using GC-MS-based metabolomics (Table 1). These include starter ingredient (*Koji,* for the preparation of *koji*-derived fermented products), fast-fermented bean paste (*cheonggukjang* and *soksungjang*), fermented paste (*doenjang*) and others (*douchi*, *meju*, *moromi*, soy sauce, and *tempe* or *tempeh*). Seo *et al.* (2018a) and Seo *et al.* (2018b) studied *koji* produced from soybean and a combination of cereals. Using a combination of SPME-GC-MS and GC-TOF-MS-based metabolomics, Seo *et al.* (2018a) compared the volatile organic compounds and primary metabolites in *koji* samples fermented individually with *Bacillus amyloliquefaciens* and *Aspergillus oryzae*. Through these integrated approaches, the authors concluded that the volatile profile of *koji* is largely determined by the inocula choice, which modifies the primary metabolites in *koji* substrates, subsequently influencing its aroma characteristics. Same authors also adopted GC-TOF-MS-based study to unravel the effects of varying substrates (soybean, wheat, and rice) and same inocula (*A. oryzae* and *B. amyloliquefaciens*) on metabolite compositions of *koji* (Seo *et al*., 2018b). The substrates influenced primary metabolite compositions in *koji* types with soybean greater than wheat and rice. Among the inocula choice for the *koji* types, *A. oryzae* was observed to have stimulated higher carbohydrates, lipid derivative, and organic acids levels while *B. amyloliquefaciens* produced higher amino acids levels, suggesting that the metabolomic approach showed promising applications toward production optimisation, bioprocess and quality control of *koji* products.

Some studies have demonstrated the metabolite profiling of fast-fermented soybean pastes (*cheonggukjang* and *soksungjang*) during fermentation with focus on answering varying biological questions. Baek *et al.* (2010) and Kim *et al.* (2012) investigated *cheonggukjang* inoculated with different *Bacillus* strains and metabolites changes with respect to fermentation times were revealed. On one hand, distinct patterns of amino acids, organic acids, sugars, and sugar alcohols were reported according to the fermentation period (0-72 h) whereas significant differences in pre-determined metabolite contents were dependent on the inocula strains (Baek *et al.,* 2010). Similarly, Kim *et al.* (2012) indicated that the separation of metabolites in *cheonggukjang* samples was mainly influenced by the fermentation duration (0-72 h). Using a targeted approach, Park *et al.* (2010) examined the changes in pre-defined metabolites (amino acids, fatty acids, and organic acids) of *cheonggukjang* and reported the major components (such as citric acid, tryptophan, leucine, among others) differentiated fermented samples according to fermentation duration (0-50 h). Furthermore, Oh *et al.* (2016) determined the metabolite profiles of four types of

cheonggukjang with added garlic using untargeted metabolic approach. The addition of garlic decreased the levels of four amino acids whereas high level of metabolic components with positive correlation with antioxidant activities was observed. The elucidated metabolite changes of *cheonggukjang* therefore suggested that fermentation period and additives may play important role in metabolic differences of fermented foods.

Studies on *soksungjang* have compared the metabolite profile of buckwheat *soksungjang* samples (BSs) obtained via traditional and commercially modified methods or inoculated with multiple microbial starters (Park *et al.,* 2017; Park *et al.,* 2019). Based on the fermentation type and fermentation period, accelerated changes and differences in the volatile compounds of commercial BSs were reported compared to the traditional type (Park *et al.,* 2017). In a follow up study, variations in the volatile profile of BSs (mainly, acids, benzenes, and esters) were found to depend on the microbial starter combinations as well as fermentation periods (Park *et al.,* 2019). These findings may provide insight for optimising the fermentation process of BS.

Fermented soybean pastes with longer fermentation period have also been investigated using metabolomic approach. Jeong *et al.* (2017) determined the effect of bacterial species on the volatile compound profiles of fermented sterilised soybeans. In comparison with uncultured fermented soybean, *Enterococcus faecium* and *Tetragenococcus halophilus* fermented beans produced similar volatile compound profiles whereas *Bacillus licheniformis* and *Staphylococcus succinus* induced vital volatile compounds (2,3,5,6-tetramethylpyrazine, 3-methylbutyl acetate, and phenylmethanol) that differentiated the fermented soybean flavour. The latter starter candidates, *B. licheniformis* and *S. succinus* along with *T. halophilus* have also been used as different starter combinations to ferment sterilised soybeans with the addition of NaCl, and corresponding effects on volatile compounds profiles were determined thereof (Jeong *et al.,* 2019). At a good cell growth, *B. licheniformis* and *S. succinus* significantly contribute to the production of a specific volatile compound profile. However, the concentration increases in NaCl from 1.5 to 14% in the mixed culture showed dominance of the starters were shifted to *T. halophilus*. Soybean culture containing *S. succinus* and 7% NaCl was reported to produce decisive volatile compounds; 3-methylbutan-1-ol and octan-3-one. Moreover, the authors demonstrated that the flavor profile and microbial dominance of the soybean culture can be manipulated by the inclusion of NaCl (Jeong *et al.,* 2019).

Park and Kim (2020) investigated the differences in metabolite profiles of fermented soybeans induced by various microbial starters. Using the PLS-DA analysis for volatile metabolite profiles. Their study revealed that the fungi group was evidently discriminated from the bacteria group. Regarding metabolic pathways, the formation of fatty acids-derived volatiles was higher in the bacteria group while that of branched-chain aliphatic alcohols and esters increased in the fungi group. As such, depicting the microbial-specific role and impact on the metabolites produced during soybean fermentation. In another study by Sun *et al.* (2019), differentially induced metabolite profiles in soybean pastes were shown by two strains. The result obtained revealed that α-ketoglutaric acid-derived amino acid and oxaloacetate-derived type synthesis are predominant in *Penicillium glabrum* GQ1-3 and *Aspergillus oryzae* HGPA20, respectively. They concluded that the different pathways of amino acid synthesis lead to the distinct nutrients and umami substances in the fermented soybean pastes.

Doenjang is also a type of fermented soybean paste. Using commercial and traditional procedures for the preparation of *doenjang*, Jo *et al*. (2011) used GC-MS metabolomics to describe differences in the volatile compounds of commercial samples. Metabolites reported were ethanol, ethyl esters, and maltol while acids, carbonyls, furans, phenols, and pyrazines were found in traditional doenjang samples. Similarly, Lee *et al*. (2017) investigated the metabolite profiles of *doenjang* samples produced via industrial and modified industrial manufacturing processes with specific microbial inocula and reported that the metabolites quantified showed distinct patterns with respect to fermentation processes. In addition, they revealed that the metabolism of amino acids, fatty acids, and sugars were associated with *Zygosaccharomyces rouxii*, *Bacillus velezensis* and *A. oryzae*, respectively. Jeong *et al*. (2020) used *Enterococcus faecium*, *Staphylococcus succinus*, and *A. oryzae* in the production of *doenjang* and investigated the effect of bacterial starters on flavour production during fermentation process. GC-MS metabolomics revealed that flavour development in the *doenjang* samples cultured with bacterial starters were related to 2-methylbutanoic acid, 3-methylbutanoic acid and acetic acid, as the important volatile compounds. It could therefore be suggested that *doenjang* composition and final quality may be significantly influenced by the microbial population involved in manufacturing process. Likewise, Lee *et al*. (2014) conducted a comprehensive GC-MS metabolomic profiling of *doenjang* at different steps of its industrial processing. At the following steps; *meju* fermentation, brining, and *doenjang* aging, increases in some primary metabolites such as amino acids, fatty acids, and monosaccharides were suggested according to their metabolic pathways.

Li *et al*. (2019) evaluated the dynamics of metabolome of *douchi* fermentation by using untargeted GC-TOF-MS-based metabolomics. The metabolites of different fermentation times were compared using MVDA techniques (PCA and OPLS-DA). The authors reported separation of the metabolites within 15 days of fermentation, with no discrimination reported after 15 days of fermentation. Results obtained also suggested that fermentation of *douchi* was finished in 15 days, with levels of metabolites such as alanine, lysine, putrescine, myo-inositol and L-malic acid varying significantly throughout the processes. Using GC-MS-based metabolomics, Kadar *et al*. (2018) effectively classified *tempe* from seven cities in Indonesia and three different production processes in Indonesia. The study also identified metabolites that are associated with the differently processed *tempe* to include amino acids, organic acids, other compounds, sugars, and some unknown compounds. According to these authors, the differences in the metabolome of various *tempe* samples may be due to different ambient temperature of each region that affected microbial communities during *tempe* fermentation.

Given the above studies, the metabolomic approach based on GC-MS has enhanced the comprehensive understanding of the distinct characteristics of CLFFs. These are relatively to the effect of several factors including substrate composition, additives (NaCl and garlic), fermentation time, fermentative microbes, and fermentation type. The fundamental knowledge of these factors will enable the prediction of optimal fermentation time, optimal flour/microbes/microbial strain combination, monitor fermentation process as well as determine volatile metabolite markers/SDMs, among others. All as a whole may suggest better food quality for the development of novel CLFFs with uniform characteristics. Furthermore, valuable insights of metabolomics data could accelerate the acceptance and rational promotion of promising innovative products such as yeast-free leavened doughs, gluten free bread, and nutrient-rich food waste (BSG).

4 Challenges of metabolomics in CLFF studies in a developing world

One of the main challenges related to the development, realisation and establishment of metabolomics research in CLFF field, particularly in the developing world, is the cost involved in such (fundamental and translational) studies. For instance, a modern high-resolution GC-TOF-MS system may cost above US$800,000, and such amount does not include the running and maintenance costs. It is worth noting also that in addition to GC-MS systems, other analytical platforms, such as LC-MS may be required for better coverage of the chemical space of CLFFs. The few metabolomics studies to have utilised other analytical systems (Table 1) reflect the low

possibility of adopting a multi-platform metabolomics approach. The latter has also its own challenges related to data fusion and integration, in additional to cost involved. Fund accessibility is significantly limited in the developing world; and subsequently, national and private research (academic and non-academic) institutions often have no means to afford these analytical platforms required for metabolomics studies.

In addition to cost, another bottleneck in the establishment of metabolomics research in the developing world is training (in metabolomics) and awareness (or publicity) of the metabolomics field and its applicability potentials in food science and technology. In Africa, for instance, metabolomics or fundamentals of this multidisciplinary omics science are not yet incorporated in academic institutions curricula. This limits the pool of skilled scientists that would adventure into metabolomics-related philosophy and research; affecting subsequently the popularity of the field. Furthermore, much effort is still needed by scientists working in the food metabolomics field at propagating the application of metabolomics in the food industry, particularly in the developing world. It is in the view of the authors that, with a growing metabolomics community in South Africa (https://www.metabolomics-sa.co.za/), the field of metabolomics will gradually expand and mature in Africa; but more efforts are still needed.

5 Future direction and conclusion

The emergent application of metabolomics in food science, particularly in food processes, is still necessary to mature and be integrated with other approaches in the field. Much more still needs to be done by researchers in this field and the challenges highlighted in Section 4.0 also re-emphasises the need for more collaboration among food scientists/technologists and the metabolomics community. The recent studies highlighted in this review also suggests that the application of GC-MS-based metabolomics for CLFFs is undeniably gradually gaining momentum and providing insights in resolving the chemical diversity and metabolic information involved in fermentation (Figure 2). Since metabolite concentration varies from one source to another and some metabolites with lower concentrations may have important regulatory effects, the detection of these trace-level metabolites presents a challenge for MS-based metabolomics. Therefore, advancements around the combined use of multiple separation, extraction methods and the analytical capabilities of GC-MS instrumentation such as GC-MS/MS, GC×GC-ToFMS and recently LC×GC combo all with improved sensitivities would also provide a better overview of the constituents in fermented foods. Considering the plethora of fermented foods around the

world, much more still needs to be done in bridging this knowledge gap. However, a proper understanding of metabolomics workflow is still essential and in its infancy for the study of CLFFs. This implies the need for adequate metabolomics training and more frequent use in the food science and technology community. The weaknesses encountered in one metabolomics approach is usually the respective strengths of the other, hence the integration of metabolomics approaches with other omics methods (meta-analysis) will significantly contribute to a more detailed and comprehensive characterization of the fermentation processes under investigation and leading to better understanding and more discoveries.

Acknowledgements

The University of Johannesburg (UJ) Global Excellence and Stature (GES) 4.0 Catalytic Initiative Grant, UJ Research Committee Grant and the National Research Foundation (NRF) of South Africa Thuthuka funding (Grant no: 121826) are duly acknowledged.

Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Ethical Guidelines

Ethical approval was not required for this research.

Conflict of interest

The authors declare no conflict of interest.

References

Adahchour, M., Brandt, M., Baier, H.U., Vreuls, R.J., Batenburg, A.M., & Brinkman, U.A. (2005). Comprehensive two dimensional gas chromatography coupled to a rapid-scanning quadrupole mass spectrometer: Principles and applications. *Journal of Chromatography A*, 1067, 245–254.

*Adamski, J. (2020). Introduction to metabolomics. In: J. Adamski (Ed.). Metabolomics for Biomedical Research (pp. 1–15). Netherlands: Elsevier.

*In this work, the author showed detailed information on introduction to metabolomics.

*Adebo, O.A., Njobeh, P.B., Adebiyi, J.A., Gbashi, S., & Kayitesi, E. (2017a). Food metabolomics (Foodomics), a new frontier in food analysis and its potential in understanding

fermented foods. In: M.C. Hueda (Ed.). Functional Food - Improve Health through Adequate Food (pp. 211–234). Croatia: InTech.

*In this chapter, the authors provided information regarding the application of food metabolomics to fermented foods in published works and identified the analytical approach prospects.

Adebo, O.A., Njobeh, P.B., Adebiyi, J.A., Gbashi, S., Phoku, J.Z., & Kayitesi, E. (2017b) Fermented pulse-based foods in developing nations as sources of functional foods. In: M.C. Hueda (Ed). Functional Food–Improve Health through Adequate Food (pp. 77–109). Croatia: InTech.

Adebo, O.A., Kayitesi, E., Tugizimana, F., & Njobeh, P.B. (2019). Differential metabolic signatures in naturally and lactic acid bacteria (LAB) fermented *ting* (a Southern African food) with different tannin content, as revealed by gas chromatography mass spectrometry (GC–MS)-based metabolomics. *Food Research International*, 121, 326–335.

Adebo, O.A. (2020). African sorghum-based fermented foods: Past, current and future prospects. *Nutrients*, 12, 1111.

Adebo, O.A., & Medina-Meza, I.E. (2020). Impact of fermentation on the phenolic compounds and antioxidant activity of whole cereal grains: A mini review. *Molecules*, 25, 927.

Baek, J.G., Shim, S.-M., Kwon, D.Y., Choi, H.-K., Lee, C.H., & Kim, Y.-S. (2010). Metabolite profiling of *cheonggukjang*, a fermented soybean paste, inoculated with various *Bacillus* strains during fermentation. *Bioscience, Biotechnology, and Biochemistry*, 74, 1860–1868.

Balestra, F., Laghi, L., Saa, D.T., Gianotti, A., Rocculi, P., & Pinnavaia, G. (2015). Physico-chemical and metabolomic characterization of KAMUT Khorasan and durum wheat fermented dough. *Food Chemistry*, 187, 451–459.

Barnes, S. (2020). Overview of experimental methods and study design in metabolomics, and statistical and pathway considerations. In: S Li (Ed.). Computational Methods and Data Analysis for Metabolomics (pp. 1–10). USA: Springer.

Beale, D.J., Pinu, R.P., Kouremenos, K.A., Poojary, M.M., Narayana, V.K., Boughton, B.A., Kanojia, K., Dayalan, S., Jones, O.A.H., & Dias, D.A. (2018). Review of recent developments in GC–MS approachesto metabolomics-based research. *Metabolomics*, 14, 152.

Blandino, A., Al-Aseeri, M.E., Pandiella, S.S., Cantero, D., & Webb. C. (2003). Cereal-based fermented foods and beverages. *Food Research International*, 36, 527–543.

Bouhifd, M., Beger, F., Flynn, T., Guo, L., Harris, G., Hogberg, H., Kaddurah-Daouk, R., Kamp, H., Kleensang, A., Maertens, A., Odwin-DaCosta, S., Pamies, D., Robertson, D., Smirnova, L., Sun, J., Zhao, L., & Hartung, T. (2015). Quality assurance of metabolomics. *Altex*, 32, 319–326.

Broadhurst, D., Goodacre, R., Reinke, S.N., Kuligowski, J., Wilson, I.D., Lewis, M.R., & Dunn, W.B. (2018). Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics*, 14, 72.

Cajka, T. (2013). Gas chromatography-time-of-flight mass spectrometry in food and environmental analysis. *Comprehensive Analytical Chemistry,* 61, 271–302.

Cevallos-Cevallos, J.M., Reyes-De-Corcuera, J.I., Etxeberria, E., Danyluk, M.D., & Rodrick, G.E. (2009). Metabolomic analysis in food science: A review. *Trends in Food Science and Technology*, 20, 557–566.

Cifuentes, A. (2009). Food analysis and foodomics. *Journal of Chromatography A*, 1216, 7109.

Cooray, S.T., & Chen, W.N. (2018). Valorization of brewer's spent grain using fungi solid-state fermentation to enhance nutritional value. *Journal of Functional Foods*, 42, 85–94.

Dettmer, K., Aronov, P.A., & Hammock, B.D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*, 26, 51–78.

Dunn, W.B., & Hankemeier, T. (2013). Mass spectrometry and metabolomics: Past, present and future. *Metabolomics*, 9, S1–S3.

Dunn, W.B., Wilson, I.D., Nicholls, A.W., & Broadhurst, D. (2012). The importance of experimental design and qc samples in largescale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis*, 4, 2249–2264.

Feng, Z., Ding, C., Li, W., Wang, D., & Cui, D. (2020). Applications of metabolomics in the research of soybean plant under abiotic stress. *Food Chemistry*, 310, 125914.

Ferri, M., Serrazanetti, D.I., Tassoni, A., Baldissarri, M., & Gianotti, A. (2016). Improving the functional and sensorial profile of cereal-based fermented foods by selecting *Lactobacillus plantarum* strains via a metabolomics approach. *Food Research International*, 89, 1095–1105.

*Fiehn, O. (2017). Metabolomics by gas chromatography-mass spectrometry: the combination of targeted and untargeted profiling. *Current Protocols in Molecular Biology*, 114, 1–30.

*The study provided an overview of GC-MS-metabolomics, highlighting some of the challenges and advantages of this metabolomic approach.

Fjeldsted, J.C. (2016). Advances in time-of-flight mass spectrometry. *Comprehensive Analytical Chemistry,* 71, 19–49.

Gbashi, S., Adebo, O.A., Piater, L., Madala, N.E., & Njobeh, P.B. (2017). Subcritical water extraction of biological materials. *Separation & Purification Reviews*, 46, 21–34.

Groger, T.M., Kafer, U., & Zimmermann, R. (2020). Gas chromatography in combination with fast high-resolution time-of-flight mass spectrometry: Technical overview and perspectives for data visualization. *Trends in Analytical Chemistry*, 122, 115677.

Gu, Q., David, F., Lynen, F., Rumpel, K., Dugardeyn, J., Van Der Straeten, D., Xu, G., & Sandra, P. (2011). Evaluation of automated sample preparation, retention time locked gas chromatography-mass spectrometry and data analysis methods for the metabolomic study of arabidopsis species. *Journal of Chromatography A*, 1218, 3247–3254.

Hyeon, H., Min, C.W., Moon, K., Cha, J., Gupta, R., Park, S.U., Kim, S.T., & Kim, J.K. (2020). Metabolic profiling-based evaluation of the fermentative behavior of *Aspergillus oryzae* and *Bacillus subtilis* for soybean residues treated at different temperatures. *Foods*, 9, 117.

Iijima, Y. (2014). Recent advances in the application of metabolomics to studies of biogenic volatile organic compounds (BVOC) produced by plant. *Metabolites*, 4, 699–721.

Jeong, D.-W., Heo, S., Lee, B., Lee, H., Jeong, K., Her, J.-Y., Lee, K.-G., & Lee, J.-H. (2017). Effects of the predominant bacteria from *meju* and *doenjang* on the production of volatile compounds during soybean fermentation. *International Journal of Food Microbiology*, 262, 8–13.

Jeong, D.-W., Lee, H., Jeong, K., Kim, C.-T., Shim, S.-T., & Lee, J.-H. (2019). Effects of starter candidates and NaCl on the production of volatile compounds during soybean fermentation. *Journal of Microbiology and Biotechnology*, 29, 191–199.

Jeong, D.-W., Jeong, K., Lee, H., Kim, C.-T., Heo, S., Oh, Y., Heo, G., & Lee, J.-H. (2020). Effects of *Enterococcus faecium* and *Staphylococcus succinus* starters on the production of volatile compounds during doenjang fermentation. *LWT - Food Science and Technology*, 122, 108996.

Jo, Y.-J., Cho, I.H., Song, C.K., Shin, H.W., & Kim, Y.-S. (2011). Comparison of fermented soybean paste (*doenjang*) prepared by different methods based on profiling of volatile compounds. *Journal of Food Science*, 76, C368-C379.

Kadar, A.D., Aditiawati, P., Astawan, M., Putri, S.P., & Fukusaki, E. (2018). Gas chromatography coupled with mass spectrometry-based metabolomics for the classification of *tempe* from

different regions and production processes in Indonesia. *Journal of Bioscience and Bioengineering*, 3, 411–416.

Kewuyemi, Y.O., Njobeh, P.B., Kayitesi, E., Adebiyi, J.A., Oyedeji, A.B., Adefisoye, M.A., & Adebo, O.A. (2020a). Metabolite profile of whole grain *ting* (a Southern African fermented product) obtained using two strains of *Lactobacillus fermentum*. *Journal of Cereal Science*, 95, 103042.

Kewuyemi, Y.O., Hema, K., Chinma, C.E., & Adebo, O.A. (2020b). Fermented edible insects for promoting food security in Africa. *Insects*, 11, 283.

Kim, J., Choi, J.N., John, K.M.M., Kusano, M., Oikawa, A., Saito, K., & Lee, C.H. (2012). GC−TOF-MS- and CE−TOF-MS-Based metabolic profiling of *Cheonggukjang* (fast-fermented bean paste) during fermentation and its correlation with metabolic pathways. *Journal of Agricultural and Food Chemistry*, 60, 9746–9753.

Kohajdova, Z., & Karovicova, J. (2007). Fermentation of cereals for specific purpose. *Journal of Food and Nutrition Research*, 46, 51–57.

Kum, S.-J., Yang, S.-O., Lee, S. M., Chang, P.-S., Choi, Y. H., Lee, J. J., Hurh, B. S., & Kim, Y.-S. (2015). Effects of *Aspergillus* species inoculation and their enzymatic activities on the formation of volatile components in fermented soybean paste (*doenjang*). *Journal of Agricultural and Food Chemistry*, 63, 1401–1418.

Kwon, Y.S., Lee, S., Lee, S.H., Kim, H.J., & Lee, C.H. (2019). Comparative evaluation of six traditional fermented soybean products in East Asia: A metabolomics approach. *Metabolites*, 9, 183.

Lee, D.E., Lee, S., Jang, E.S., Shin, H.W., Moon, B.S., & Lee, C.H. (2016). Metabolomic profiles of *Aspergillus oryzae* and *Bacillus amyloliquefaciens* during rice *Koji* fermentation. *Molecules*, 21, 773.

Lee, S.M., Jung, J.H., Seo, J-A., & Kim, Y-S. (2018). Bioinformation of volatile and nonvolatile metabolites by *Saccharomycopsis fibuligera* KJJ81 cultivated under different conditions—carbon sources and cultivation times. *Molecules*, 23, 2762.

Lee, S.M., Kim, S.-B., & Kim, Y.-S. (2019). Determination of key volatile compounds related to long-term fermentation of soy sauce. *Journal of Food Science*, 84, 2758–2776.

Lee, S.Y., Kim, H.Y., Lee, S., Lee, J.M., Muthaiya, M.J., Kim, B.S., Oh, J.Y., Song, C.K., Jeon, E.J., Ryu, H.S., & Lee, C.H. (2012). Mass spectrometry-based metabolite profiling and bacterial

diversity characterization of Korean traditional *Meju* during fermentation. *Journal of Microbiology and Biotechnology*, 22, 1523–1531.

Lee, D.Y., Park, J-J., Barupal, D.K., & Fiehn, O. (2012). System response of metabolic networks in *Chlamydomonas reinhardtii* to total available ammonium. *Molecular & Cellular Proteomics*, 11, 973–988.

Lee, S.Y., Lee, S., Lee, S., Oh, J.Y., Jeon, E.J., Ryu, H.S., & Lee, C.H. (2014). Primary and secondary metabolite profiling of *doenjang*, a fermented soybean paste during industrial processing. *Food Chemistry*, 165, 157–166.

Lee, S., Lee, S., Singh, D., Oh, J. Y., Jeon, E. J., Ryu, H.S., Lee, D.W., Kim, B.S., & Lee, C. H. (2017). Comparative evaluation of microbial diversity and metabolite profiles in *doenjang*, a fermented soybean paste, during the two different industrial manufacturing processes. *Food Chemistry*, 221, 1578–1586.

León, C., Cifuentes, A., & Valdés, A. (2018). Foodomics applications. *Comprehensive Analytical Chemistry*, 82, 643–685.

Lesur, A., Gallien, S., & Domon, B. (2020). Hyphenation of fast liquid chromatography with high-resolution mass spectrometry for quantitative proteomics analyses. *Trends in Analytical Chemistry*, 84, 114–150.

Li, P., Tang, H., Shi, C., Xie, Y., Zhou, H., Xia, B., Zhang, C., Chen, L., & Jiang, L. (2019). Untargeted metabolomics analysis of *Mucor racemosus douchi* fermentation process by gas chromatography with time-of-flight mass spectrometry. *Food Science and Nutrition*, 7, 1865–1874.

Li, S., Tian, Y., Jiang, P., Lin, Y., Liu, S., & Yang, H. (2020). Recent advances in the application of metabolomics for food safety control and food quality analyses. *Critical Reviews in Food Science and Nutrition*, 1–22.

Lorenzo, M., & Pico, Y. (2017). Gas chromatography and mass spectroscopy techniques for the detection of chemical contaminants and residues in foods. In: D. Schrenk, & A. Cartus (2nd Ed.), Chemical Contaminants and Residues in Food (pp 15–50). United Kingdom: Elsevier.

Lu, H., Liang, Y., Dunn, W.B., Shen, H., &Kell, D.B. (2008). Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. *TrAC Trends in Analytical Chemistry*, 27, 215–227.

Mozzi, F., Ortiz, M. E., Bleckwedel, J., De Vuyst, L., & Pescuma, M. (2013). Metabolomics as a tool for the comprehensive understanding of fermented and functional foods with lactic acid bacteria. *Food Research International*, 54, 1152–1161.

MSI (Metabolomics Standards Initiative) (2005). The Metabolomics Standards Initiative (MSI) and Core Information for Metabolomics Reporting (CIMR). http://cosmos-fp7.eu/msi.html (Accessed 20/07/2020).

Mu, Y., Su, W., Yu, X-T., Mu, Y-C., Jinag, L., & Wang, H-L. (2019). Untargeted metabolomics based on GC-TOF-MS reveals the optimal pre-fermentation time for black glutinous rice wine. *International Journal of Food Properties*, 22, 2033–2046.

Nissen, L., Rollini, M., Picozzi, C., Musatti, A., Foschino, R., & Gianotti, A. (2020a). Yeast-free doughs by *Zymomonas mobilis*: evaluation of technological and fermentation performances by using a metabolomic approach. *Microorganisms*, 8, 792.

Nissen, L., Bordoni, A., & Gianotti, A. (2020b). Shift of volatile organic compounds (VOCs) in gluten-Free hemp-enriched sourdough bread: A metabolomic approach. *Nutrients*, 12, 1050.

Oghbaei, M., & Prakash, J. (2016). Effect of primary processing of cereals and legumes on its nutritional quality: A comprehensive review. *Cogent Food & Agriculture*, 2, 1136015.

Oh, D.-G., Jang, Y. K., Woo, J. E., Kim, J.-S., & Lee, C. H. (2016). Metabolomics reveals the effect of garlic on antioxidant- and protease-activities during *cheonggukjang* (fermented soybean paste) fermentation. *Food Research International*, 82, 86–94.

Park, M. K., Cho, I. H., Lee, S., Choi, H.-K., Kwon, D.-Y., & Kim, Y.-S. (2010). Metabolite profiling of *cheonggukjang*, a fermented soybean paste, during fermentation by gas chromatography-mass spectrometry and principal component analysis. *Food Chemistry*, 122, 1313–1319.

Park, M-K., Choi, H-S., Kim, Y-S., & Cho, I. H. (2017). Change in profiles of volatile compounds from two types of *Fagopyrum esculentum* (buckwheat) *soksungjang* during fermentation. *Food Science and Biotechnology*, 26, 871–882.

*Park, M. K., & Kim, Y-S. (2019). Distinctive formation of volatile compounds in fermented rice inoculated by different molds, yeasts, and lactic acid bacteria. *Molecules*, 24, 2123.

*In this work, the authors investigated diverse approaches to compare the metabolic characteristics and identified microbe-specific metabolites in fermented rice by diverse fermentative microbes.

To compare the volatile compound profiles of fermented rice derived using different fermentative microbes

Park, M. K., Choi, H-S., Kim, Y-S., & Cho, I. H. (2019). Comparison of volatile profiles in *Fagopyrum esculentum* (buckwheat) *soksungjang* prepared with different starter cultures during fermentation. *Food Science and Biotechnology*, 28, 1037–1045.

Park, M. K., & Kim, Y.-S. (2020). Comparative metabolic expressions of fermented soybeans according to different microbial starters. *Food Chemistry*, 305, 125461.

Patil, S., Brennan, M. A., Mason, S. L., & Brenna, C. S. (2016). The potential of combining cereals and legumes in the manufacture of extruded products for a healthy lifestyle. *EC Nutrition*, 5.2, 1120–1127.

Pavagadhi, S., & Swarup, S. (2020). Metabolomics for evaluating flavor-associated metabolites in plant-based products. *Metabolites*, 10, 197.

Qualley, A. V., & Dudareva, N. (2009). Metabolomics of plant volatiles. In: D. Belostotsky (Eds.) Plant Systems Biology (pp 329–343). Methods in Molecular Biology™ (Methods and Protocols), vol. 553. London: Humana Press.

Raghuvanshi, R., Grayson, A. G., Schena, S., Amanze, O., Suwintono, K., & Quinn, R. A. (2019). Microbial transformations of organically fermented foods. *Metabolites*, 9, 165.

Roberts, L. D., Souza, A. L., Gerszten, R. E., & Clish, C. B. (2012). Targeted metabolomics. *Current Protocol in Molecular Biology*, 98, 1–24.

Saa, D. L. T., Nissen, L., & Gianotti, A. (2019). Metabolomic approach to study the impact of flour type and fermentation process on volatile profile of bakery products. *Food Research International*, 119, 510–516.

Seo, S.-H., Park, S.-E., Yoo, S.-A., Lee, K. I., Na, C.-S., & Son, H.-S. (2016). Metabolite profiling of *Makgeolli* for the understanding of yeast fermentation characteristics during fermentation and aging. *Process Biochemistry*, 51, 1363–1373.

Seo, S. H., Lee, S., Singh, D., Park, M. K., Kim, Y-S., Shin, H. W., Cho, S. A., & Lee, C. H. (2018a). Evaluating the headspace volatolome, primary metabolites, and aroma characteristics of *koji* fermented with *Bacillus amyloliquefaciens* and *Aspergillus oryzae*. *Journal of Microbiology and Biotechnology*, 28, 1260–1269.

Seo, S. H., Lee, S., Singh, D., Shin, H. W., Cho, S. A., & Lee, C. H. (2018b). Untargeted metabolite profiling for *koji*-fermentative bioprocess unravels the effects of varying substrate types and microbial inocula. *Food Chemistry*, 266, 161–169.

Simmler, C., Kulakowski, D., Lankin, D.C., McAlpine, J.B., Chen, S-N., & Pauli, G.F. Holistic analysis enhances the description of metabolic complexity in dietary natural products. *Advances in Nutrition*, 7, 179-189.

Singh, D., Lee, S., & Lee, C. H. (2017). Metabolomics for empirical delineation of the traditional Korean fermented foods and beverages. *Trends in Food Science and Technology*, 61, 103–115.

Song, X., Jing, S., Ma, C., Song, T., Wu, J., Zhao, Q., Zheng, F., Zhao, M., & Chen, F. (2020). Untargeted and targeted metabolomics strategy for the classification of strong aroma-type *baijiu* (liquor) according to geographical origin using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry. *Food Chemistry*, 314, 126098.

Sun, X., Lyu, G., Luan, Y., Yang, H., & Zhao, Z. (2019). Metabolomic study of the soybean pastes fermented by the single species *Penicillium glabrum* GQ1-3 and *Aspergillus oryzae* HGPA20. *Food Chemistry*, 295, 622–629.

Ten-Doménech, I. T., Ramos-Garcia, V., Piñeiro-Ramos, J. D., Gormaz, M., Parra-Llorca, A., Vento, M., Kuligowski, J., & Quintás, G. (2020). Current practice in untargeted human milk metabolomics. *Metabolites*, 10, 43.

Tugizimana F, Paiter L, & Dubery I. (2013). Plant metabolomics: a new frontier in phytochemical analysis. *South African Journal of Science*, 109, 1–11.

Valdés, A., Cifuentes, A., & León, C. (2017). Foodomics evaluation of bioactive compounds in foods. *Trends in Analytical Chemistry*, 96, 2–13.

Verbeke, K. A., Boobis, A. R., Chiodini, A., Edwards, C. A., Franck, A., Kleerebezem, M., Nauta, A., Raes, J., van Tol, E. A. F., Tuohy, K. M., &ILSI Europe Prebiotics Task Force Expert Group. (2015). Towards microbial fermentation metabolites as markers for health benefits of prebiotics. *Nutrition Research Reviews*, 28, 42–66.

Verni, M., Verardo, V., & Rizzello, C. G. (2019). How fermentation affects the antioxidant properties of cereals and legumes. *Foods*, *8*, 362.

Villas-Bôas, S.G., Mas, S., Akesson, M., Smedsgaard, J., & Nielsen, J.(2005). Mass spectrometry in metabolome analysis. *Mass Spectrometry Reviews*, 24, 613–646

Wang, L., Naser, F. J., Spalding, J. L., & Patti, G. J. (2019). A protocol to compare methods for untargeted metabolomics. *Methods in Molecular Biology*, 1862, 1–15.

Wang, N., Chen, S., & Zhou, Z. (2020). Age-dependent characterization of volatile organic compounds and age discrimination in Chinese rice wine using an untargeted GC/MS-based metabolomic approach. *Food Chemistry*, 325, 126900.

Weckx, S., Van der Meulen, R., Maes, D., Scheirlinck, I., Huys, G., Vandamme, P., & De Vuyst, L. (2010). Lactic acid bacteria community dynamics and metabolite production of rye sourdough fermentations share characteristics of wheat and spelt sourdough fermentations. *Food Microbiology*, 27, 1000–1008.

Yin, H., Dong, J.-J., Yu, J.-H., Zhong, C., Liu, L.-P., Liu, M., Yang, P.-F., Hao, J.-G., & Jia, S.-R. (2017). Intracellular metabolite profiling of industrial yeast and the synthesis of flavour compounds in beer. *Journal of the Institute of Brewing*, 123, 328–336.

Zheng, J., Wu, C.-D., Huang, J., Zhou, R.-Q., & Liao, X.-P. (2013). Analysis of volatile compounds in Chinese soy sauces *moromi* cultured by different fermentation processes. *Food Science and Biotechnology*, 22, 605–612.

Table 1: Some recent studies on GC-MS based metabolomics of cereal and legume-based fermented foods

Produce	Fermented food product	Approach	Derivatization	Analytical platform	Data processing platform	Data mining tool	Metabolite groups reported	Biological question/primary aim of study	Reference
Cereal based FFs									
Beverages (alcoholic and non-alcoholic)									
Black glutinous rice	Black glutinous rice wine	Untargeted	Methoxy-amination and addition of BSTFA	GC-TOF-MS	ChromaTOF 4.3X, LECO-Fiehn Rtx5 database	PCA, OPLS-DA, Metabo -Analyst	Amino acids, organic acids, fatty acids, phenolic acids, sugars, and sugar alcohols	To investigate the relationship between metabolites and fermentation time	Mu *et al*. (2019)
Cereals	*Baijiu*	Untargeted and targeted*	None	GC×GC-TOFMS	ChromaTOF	PCA, PLS-DA	Alcohol, esters, furan, ketones, organic acid, pyrazine, and sulphur containing compounds	To connect the geographical origin to the volatile composition of *baijiu* samples	Song *et al.* (2020)
Rice	Chinese rice wine	Untargeted	None	HS-SPME-GC-MS SPE-GC-MS	Chemstation XCMS	PLSR	Acids, alcohols, aldehydes, aromatics, ethyl esters, furans, ketones, lactones, pyrazines, sulfides, phenols and their derivatives	To characterize VOCs during Chinese rice wine aging	Wang *et al.* (2020)
Rice	*Makgeolli*	Untargeted	Methoxy-amination and addition of MSTFA	GC-MS	Chemstation	PCA, OPLS-DA	Amino acids, organic acids, polyols, and sugars	To investigate the metabolic changes in *Makgeolli* during alcoholic fermentation and aging	Seo *et al.* (2016)
NR	Lager beer	Untargeted	Methoxy-amination and addition of MSTFA	GC-MS	Agilent MSD	PCA, PLS	Alcohols, amino acids, fatty acids, organic acids, and sugars	To explore the global intracellular metabolite profiles of lager yeast during brewing fermentation	Yin *et al.* (2017)
Dough, gruel, bread									
Wheat	Yeast-free doughs	Targeted	None	SPME-GC-MS	NR	PCA, K-means clustering	Organic acids, esters, aldehydes, and ketones	To profile the volatile compounds in yeast-free dough	Nissen *et al.* (2020a)
Maize, rice, and	Gluten-free bread	Untargeted	None	SPME-GC-	NR	PCA, K-	Alcohols, aldehydes, alkenes,	To evaluate the effect of flour	Nissen *et al.*

Produce	Fermented food product	Approach	Derivatization	Analytical platform	Data processing platform	Data mining tool	Metabolite groups reported	Biological question/primary aim of study	Reference
hemp seed				MS		means clustering	ketones, and organic acids	addition and sourdough fermentation on the production of VOCs in gluten-free bread	(2020b)
Sorghum	*Ting*	Untargeted	None	GC-HRTOF-MS	XCMS	PCA, OPLS-DA	Benzene, cyclic compounds, esters, fatty acid derivatives, ketones, organic acids, pesticide, furan,phenols, and sugar derivatives	To descriptively elucidate metabolic profiles of sorghum types and derived fermented products obtained via controlled and spontaneous fermentation	Adebo *et al.* (2019)
Wheat (durum and KAMUT khorasan)	Fermented dough	Untargeted	NR	SPME-GC-MS	Chemstation XCMS	HCA	Acids, alcohols, carbonils, esters, and hydrocarbons	To determine the optimal flour/microbial strain combinations for sourdough preparation.	Ferri *et al.* (2016)
Wheat (Durum and KAMUT khorasan)	Bread	Untargeted	None	SPME-GC-MS	Chemstation XCMS	CAP	Aldehydes, alcohols, carboxylic acids, esters, hydrocarbons, ketones, organic acids, and phenols	To investigate the metabolites profile in sourdoughs prepared from different wheat samples at varying maturation stage	Saa *et al.* (2019)
Rice	*Koji*	Untargeted	Methoxy-amination and addition of MSTFA	GC-TOF-MS	ChromaTOF MetAlign	PCA, PLS-DA, OPLS-DA	Amino acids, fatty acids, organic acids, phenolic acids, vitamins, sugars and sugar alcohols	To identify differences between the metabolites in rice *koji* inoculated with *Aspergillus oryzae* or *Bacillus amyloliquefaciens*	Lee *et al.* (2016)
Others									
Brewer's spent grain (BSG)	Fermented BSG	Untargeted	Methoxy-amination and addition of MSTFA	GC-MS	Agilent MassHunter and Agilent Mass Profiler	PCA, OPLS-DA	Amino acid, aminophenol, amino fatty acids, carbohydrates, tricarboxylic acids, and vitamin	To investigate the constituents of fermented BSG using different extraction solvent and derivatization method	Cooray and Chen (2018)
Rice	Fermented rice	Untargeted	None	SPME-GC-	NR	PLS-DA	Amino acid derivatives,	To compare the volatile	Park and Kim

Produce	Fermented food product	Approach	Derivatization	Analytical platform	Data processing platform	Data mining tool	Metabolite groups reported	Biological question/primary aim of study	Reference
				MS			benzenes, butanediol, ethanol, fatty acid and their derivatives, furans, lactone, and sulfur-containing compounds	compound profiles of fermented rice derived using different fermentative microbes	(2019)
Rice and soybean	Rice-*koji doenjang*	Untargeted	None	SPME-GC-MS	NR	PCA, PLSR	Alcohols, benzenes and benzene derivatives, carbonyls, esters, furans and furan derivatives, hydrocarbons, miscellaneous, organic acids, phenols, and sulfur-containing compounds	To investigate the volatile profiles of rice-*koji doenjang* inoculated with three types of *Aspergillus* species and fermented over a range of periods	Kum *et al.* (2015)
Legume based FFs									
Starter ingredient									
Soybean, wheat	*Koji*	Untargeted	Methoxy amination and addition of MSTFA	SPME-GC-MS, GC-TOF-MS	ChromaTOF Metalign	PCA, PLS-DA	Alcohols, aldehydes, aliphatic hydrocarbons, aromatic hydrocarbons, carboxylic acids, esters, furans, ketones, lactones, phenols, pyrazines, sulfur-containing compounds, miscellaneous compounds	To compare the VOCs and primary metabolites in *koji* samples fermented individually with *Bacillus amyloliquefaciens* and *Aspergillus oryzae*	Seo *et al.* (2018a)
Soybean, wheat, rice	*Koji*	Untargeted	Methoxy amination and addition of MSTFA	GC-TOF-MS	ChromaTOF	PCA, PLS-DA	Amino acids, fatty acids, nucleosides, phenolic acids, organic acids, sugar and sugar alcohols, vitamin	To unravel the effects of varying substrates (soybean, wheat, and rice) and inocula (*Aspergillus oryzae* and *Bacillus amyloliquefaciens*) on metabolite compositions of *koji*	Seo *et al.* (2018b)
Fast-fermented bean paste									
Soybean	*Cheonggukjang*	Targeted	BSTFA	GC-TOF-MS	NR	PCA, PLS-	Amino acids, organic acids,	To investigate the pre-	Baek *et al.*

Produce	Fermented food product	Approach	Derivatization	Analytical platform	Data processing platform	Data mining tool	Metabolite groups reported	Biological question/primary aim of study	Reference
						DA	sugars, and sugar alcohols	determined metabolite profile of *cheonggukjang* fermented with different *Bacillus* strains	(2010)
Soybean	*Cheonggukjang*	Targeted	Methoxy amination and addition of BSTFA	GC-MS	NR	PCA	Amino acids, fatty acids, and organic acids	To assess changes in pre-defined metabolites during the fermentation of *cheonggukjang*	Park *et al.* (2010)
Soybean	*Cheonggukjang*	Untargeted	Methoxy amination and addition of MSTFA	GC- and CE-TOF-MS	Databridge	PCA, PLS-DA	Alcohols, amines, amino acids, carbohydrates, inorganic acids, ketones, lactone, nucleosides, miscellaneous, organic acids, and vitamin B_3,	To investigate the metabolite changes in *cheonggukjang* as a function fermentation time and inoculated *Bacillus* strains	Kim *et al.* (2012)
Soybean	*Cheonggukjang*	untargeted	Methoxy amination and addition of MSTFA	GC-TOF-MS	ChromaTOF MetAlign	PCA, PLS-DA, OPLS-DA, Box and whisker plot analysis	Amino acid, fatty acid, organic acid, organic compound, sugar and sugar alcohol	To demonstrate the metabolite profile in four types of *Cheonggukjang*	Oh *et al.* (2016)
Soybean, buckwheat	*Soksungjang*	Untargeted	None	SPME-GC-MS	NR	PLS-DA	Alcohols, benzenes and benzene derivatives, carbonyls, esters, furans and furan derivatives, hydrocarbons, organic acids, phenols, pyrazines, sulfur-containing compounds, and miscellaneous compounds	To profile and compare the volatile compounds of buckwheat *Soksungjang* samples prepared using two different methods	Park *et al.* (2017)
Soybean, buckwheat	*Soksungjang*	Untargeted	None	SPME-GC-MS	NR	PLS-DA	Alcohols, aldehydes, aliphatic hydrocarbons, benzenes and benzene derivatives, esters, furans, ketones, lactones, organic acids, phenols,	To investigate the changes and differences in the volatile profiles of buckwheat *Soksungjang* inoculated with multiple microbial starters	Park *et al.* (2019)

Produce	Fermented food product	Approach	Derivatization	Analytical platform	Data processing platform	Data mining tool	Metabolite groups reported	Biological question/primary aim of study	Reference
							pyrazines, sulfur-containing compounds, terpenes, and miscellaneous compounds		
Fermented paste									
Soybean	Fermented soybean	Untargeted	None	SPME-GC-MS	NR	PCA	Alcohols, carbonyls, esters, furans, organic acids, pyrazines, and other compounds	To analyze the volatile compounds produced by selected starters during soybean culture	Jeong *et al.* (2017)
Soybean	Fermented soybean	Untargeted	None	SPME-GC-MS	NR	PCA	Alcohols, carbonyls, esters, furans, organic acids, and pyrazines	To investigate the effect of starter candidates and NaCl on the volatile compound profiles produced from soybean cultures	Jeong *et al.* (2019)
Soybean	Soybean paste	Untargeted	Methoxy amination and addition of BSTFA	GC-MS	ChromaTOF 4.3X, LECO-Fiehn Rtx5 database	PCA, OPLS-DA, MetaboAnalyst	Amino acids and their derivatives, organic acids, sugars and sugar alcohols, other compounds	To determine metabolites differences during the fermentation of soybean paste using two strains.	Sun *et al.* (2019)
Soybean	Fermented soybean	Untargeted	Methoxy amination and addition of MSTFA	GC-TOF-MS	ChromaTOF	PCA, OPLS-DA, HCA	Amino acids, fatty acids, organic acids, and sugars	To investigate the effect of fermentation by *Aspergillus oryzae*and *Bacillus subtilis on soybean* substrates extracted at different temperatures	Hyeon *et al.* (2020)
Soybean	Fermented soybean	Untargeted	Methoxy amination and addition of BSTFA	SPME-GC-MS GC-TOF-MS	NR	PLS-DA	Volatiles: benzenes, carbohydrates-derived, fatty acids-derived, furans, isoleucine-derived, lactones, leucine-derived, phenols, short-chain, sulfur-containing, valine-derived, and others	To investigate the differences in volatile and non-volatile metabolites induced in fermented soybeans by various microbial starters	Park and Kim (2020)

Produce	Fermented food product	Approach	Derivatization	Analytical platform	Data processing platform	Data mining tool	Metabolite groups reported	Biological question/primary aim of study	Reference
							Non-volatiles: amino acids, fatty acids, organic acids, sugar and sugar alcohols		
Soybean	*Doenjang*	Untargeted	None	SAFE-GCMS SPME-GC-MS	NR	PCA	Alcohols, benzenes and benzene derivatives, carbonyls, esters, furan and furan derivatives, hydrocarbons, phenols, pyrazines, pyrroles, organic acids, sulfur-containing compounds, and miscellaneous	To investigate the comprehensive volatile profile of traditional and commercial *doenjang* samples using two different extraction methods	Jo *et al.* (2011)
Soybean	*Doenjang*	Untargeted	Methoxy amination and addition of MSTFA	GC-TOF-MS	ChromaTOF metAlign	PCA, PLS-DA	Amino acids, fatty acids, organic acids, miscellaneous, sugar and sugar derivatives	To comprehensively investigate the metabolite profiles of *doenjang* at various steps of its industrial production process	Lee *et al.* (2014)
Soybean	*Doenjang*	Untargeted	Methoxy amination and addition of MSTFA	GC-TOF-MS	MetAlign	PCA, PLS-DA	Amino acids, fatty acids, organic acids, sugar and sugar derivatives	To contrive a correlative model twining the comparative metabolomes for *doenjang* industrial and modified industrial manufacturing processes	Lee *et al.* (2017)
Soybean	*Doenjang*	Untargeted	None	SPME-GC-MS	NR	PCA	Alcohols, carbonyls, esters, furans, organic acids, phenols, pyrazines and others	To examine the effect of *Enterococcus faecium*, *Staphylococcus Succinus*,and *Aspergillus oryzae* on flavour production in doenjang fermentation	Jeong *et al.* (2020)
Others									
Soybean	*Cheonggukjang*,	Untargeted	Methoxy	GC-TOF-MS	ChromaTOF	PCA, PLS-	Amino acids, fatty acids, organic	To demonstrate non-targeted	Kwon *et al.*

Produce	Fermented food product	Approach	Derivatization	Analytical platform	Data processing platform	Data mining tool	Metabolite groups reported	Biological question/primary aim of study	Reference
	doenjang, *doubanjiang*, *miso*, *natto*, and *tianmianjiang*		amination and addition of MSTFA		MetAlign	DA	acids, sugar and sugar derivatives, and other compounds	metabolite profiling of traditional fermented soybean products	(2019)
Soybean	*Douchi*	Untargeted	Methoxy amination and addition of BSTFA	GC-TOF-MS	ChromaTOF 4.3X, LECO-Fiehn Rtx5 database	PCA, OPLS-DA	Aldehyde, amino acids, ester, organic acids, sugars	To evaluate dynamics of the metabolome of *douchi* during fermentation	Li *et al.* (2019)
Soybean	*Meju*	Untargeted	Methoxy amination and addition of BSTFA	GC-MS/MS	Vx Capture, XCMS	PCA, PLS-DA	Amino acids, organic acids, sugars and sugar alcohols	To comprehensively examine the metabolite profile of *meju*	Lee *et al.* (2012)
Soybean	Soy sauce	Untargeted	None	SPME-GC-MS SBSE-GC-MS	NR	PCA, PLS-DA	Alcohols, aldehydes, benzene and benzene derivatives, esters, furan and furan derivatives, hydrocarbons, ketones, lactones, organic acids, phenols, pyrazines, pyrones, pyrroles, nitrogen-containing compounds, and sulfur-containing compounds	To investigate changes in profiles of soy sauce volatile compounds during long-term fermentation using SPME and SBSE	Lee *et al.* (2019)
Soybean	*Tempe*	Untargeted	Methoxy amination and addition of MSTFA	GC-MS	GCMS solution, MetAlign, AIoutput version 1.30	PCA	Amino acids, organic acids, other compounds, sugars, unknowns	To investigate the metabolic differences between different varieties of *tempe* from different regions and production processes	Kadar *et al.* (2018)
Soybean and wheat	*Moromi*	Untargeted	None	GC-MS	NR	HCA, PCA	Acids, alcohols, aldehydes, esters, ketones, miscellaneous compounds, phenols, and	To analyze volatile compounds of *moromi* prepared using different fermentation	Zheng *et al.* (2013)

Produce	Fermented food product	Approach	Derivatization	Analytical platform	Data processing platform	Data mining tool	Metabolite groups reported	Biological question/primary aim of study	Reference
							pyrazines	processes	

– used in combination with nuclear magnetic resonance (NMR); * - study later focused on marker compounds identified; BSTFA – N,OBis(trimethylsilyl)trifluoroacetamide; MSTFA – N-methyl-N-(trimethylsilyl)-trifluoroacetamide; VOCs–volatile organic compounds; SBSE–stir bar sorptive extraction; HCA – hierarchical clustering analysis, PLSR – partial least-squares regression, PLS-DA– partial least square-discriminant analysis; CAP – canonical analysis of principal coordinates; GC-HR-TOF-MS – gas chromatography-high resolution-time of flight-mass spectrometry; GC-MS – gas chromatography-mass spectrometry; GC-TOF-MS – gas chromatography-time of flight-mass spectrometry; NR – not reported; SDM – significant differentiating metabolites; HS-SPME-GC-MS – headspace solid phase micro-extraction-gas chromatography-mass spectrometry

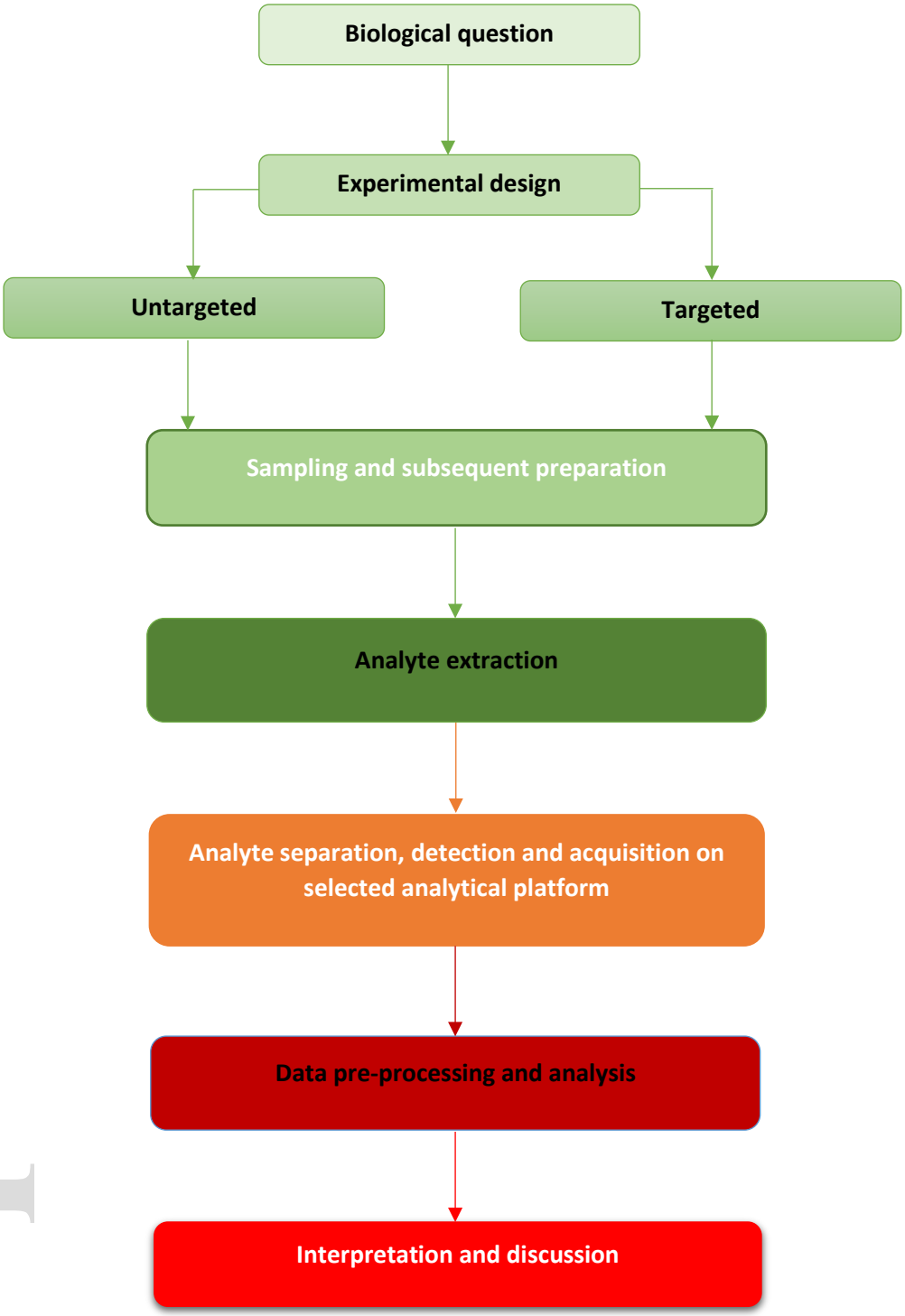

Figure 1: An overview of the metabolomics workflow

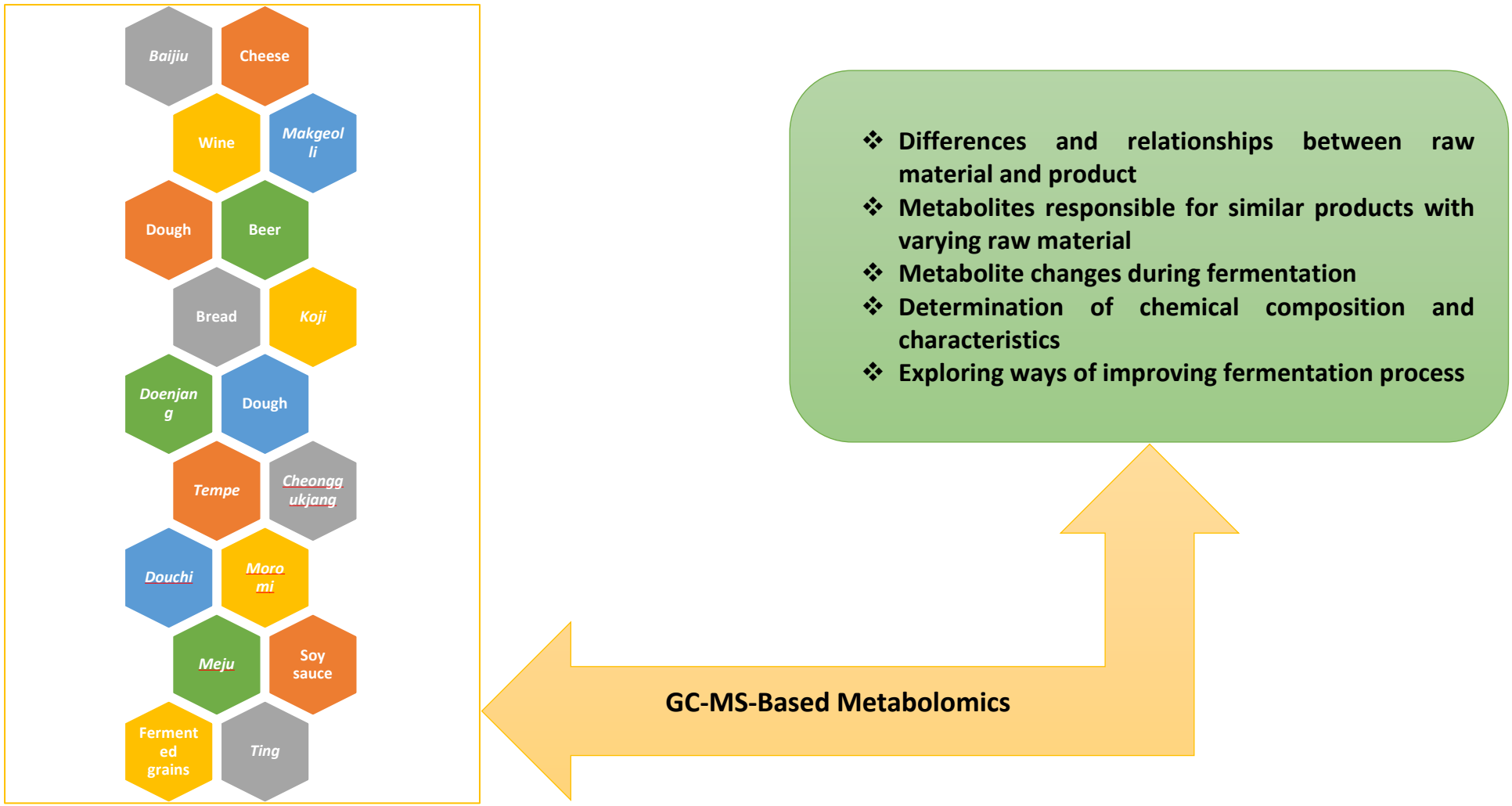

Figure 2: Metabolomic analysis of cereal and legume based fermented foods by GC-MS