



## OPEN ACCESS

EDITED BY  
Moshe Shemesh,  
Agricultural Research Organization  
(ARO), Israel

REVIEWED BY  
Ankur Naqib,  
Rush University, United States  
Jitendra Keshri,  
Medgenome Pvt. Ltd., India

\*CORRESPONDENCE  
Madhumita Barooah  
madhumita.barooah@aaau.ac.in;  
m17barooah@yahoo.co.in

SPECIALTY SECTION  
This article was submitted to  
Nutrition and Sustainable Diets,  
a section of the journal  
Frontiers in Sustainable Food Systems

RECEIVED 29 July 2022

ACCEPTED 07 November 2022

PUBLISHED 24 November 2022

CITATION  
Bora SS, Dullah S, Dey KK, Hazarika DJ,  
Sarmah U, Sharma D, Goswami G,  
Singh NR and Barooah M (2022)  
Additive-induced pH determines  
bacterial community composition and  
metabolome in traditional mustard  
seed fermented products.  
*Front. Sustain. Food Syst.* 6:1006573.  
doi: 10.3389/fsufs.2022.1006573

COPYRIGHT  
© 2022 Bora, Dullah, Dey, Hazarika,  
Sarmah, Sharma, Goswami, Singh and  
Barooah. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Additive-induced pH determines bacterial community composition and metabolome in traditional mustard seed fermented products

Sudipta Sankar Bora<sup>1</sup>, Samim Dullah<sup>2</sup>, Kuntal Kumar Dey<sup>2</sup>,  
Dibya Jyoti Hazarika<sup>2</sup>, Unmona Sarmah<sup>1</sup>, Darshana Sharma<sup>1</sup>,  
Gunajit Goswami<sup>2</sup>, Naorem Romen Singh<sup>2</sup> and  
Madhumita Barooah <sup>1,2\*</sup>

<sup>1</sup>DBT-North East Centre for Agricultural Biotechnology (DBT-NECAB), Assam Agricultural University, Jorhat, Assam, India, <sup>2</sup>Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat, Assam, India

**Introduction:** *Kahudi* and *Kharoli* are unique naturally fermented mustard seed products prepared and consumed in the northeastern region of India. The pre-fermentation processing of mustard seeds (soaking, pan-frying, mixing with alkaline or acidic additives, airtight packaging) renders a stringent fermentation environment. The metabolic activities of fermenting bacterial populations yield a myriad of glucosinolate-derived bioactive components which have not been described earlier.

**Methods:** This present study employed integrated 16S rRNA amplicon sequencing and LC-MS-based metabolomics to elucidate the bacterial diversity and metabolome of the two fermented mustard seed food products.

**Results and Discussion:** Univariate and multivariate analyses of metabolomics data revealed differential abundances of a few therapeutically-important metabolites *viz.*, sinapine, indole-3-carbinol,  $\gamma$ -linolenic acid in *Kahudi*, and metabolites *viz.*,  $\beta$ -sitosterol acetate, 3-butylene glucosinolate, erucic acid in *Kharoli*. A metagenomic investigation involving the 16S rRNA (V3–V4) amplicon sequencing showed the dominance of Firmicutes ( $99.1 \pm 0.18\%$ ) in *Kahudi*, and Firmicutes ( $79.6 \pm 1.92\%$ ) and Proteobacteria ( $20.37 \pm 1.94\%$ ) in *Kharoli*. The most abundant genera were *Bacillus* ( $88.7 \pm 1.67\%$  in *Kahudi*;  $12.5 \pm 1.75\%$  in *Kharoli*) followed by *Lysinibacillus* ( $67.1 \pm 2.37\%$  in *Kharoli*;  $10.4 \pm 1.74\%$  in *Kahudi*). Members of both these genera are well known for proteolytic and endospore-forming abilities which could have helped in colonizing and thriving in the stringent fermentation environments.

## KEYWORDS

mustard seed fermentation, LC-MS, amplicon sequencing, microbiota, metabolome

## Introduction

Rapeseed (*Brassica campestris* var. *sarson* and *toria*) and mustard (*Brassica juncea*) plants belonging to the *Brassicaceae* family are annual herbs cultivated in the vast agricultural lands of South-East Asia and the Mediterranean regions (Cartea et al., 2010; Jaime et al., 2018). The seeds of these plants are nutritionally rich in proteins (28–36%)

and dietary oils (28–32%) promoting the plants' wide acceptance and cultivation (Al-Jasass and Al-Jasser, 2012). The dietary oils extracted from the seeds are a good source of essential fatty acids, linolenic (Omega-3) and linoleic (Omega-6) acids in desirable ranges ( $11.30 \pm 6.09\%$  and  $25.31 \pm 5.74\%$  respectively) (Sharafi et al., 2015), have low amounts of saturated fatty acids, and equal proportions of mono- and poly-unsaturated fatty acids, etc. (Orsavova et al., 2015). Rapeseed/mustard oil is currently recognized as one of the healthiest edible oil options available in the market due to its rich reserves of Omega-3, antioxidants, phytosterols, tocopherols, vitamin K, etc. that exhibit cardio-protective, anti-carcinogenic and antimicrobial properties (Manchanda and Passi, 2016). Despite these benefits, remnants of many antinutritional factors such as glucosinolates (3–8% in rapeseed and 4.04% in mustard), erucic acid ( $11.35 \pm 13.83\%$ ), sinapine (about 1%), phytates (inositol hexaphosphate,  $2 \pm 4\%$ ), and tannins, etc. render it difficult to include the seed-meals of these crops into regular human and animal diets (Kumar et al., 2010; Mazumder et al., 2016). Glucosinolates (1-thiol- $\beta$ -S-D-glucosides) commonly exist as salts of potassium and are considered a precursor to many secondary antinutritional factors (ANFs) (Clarke, 2010). Hydrolysis of glucosinolates through the action of endogenous myrosinase (thioglucoside glucohydrolase) yields a myriad of several other S and/or N-containing ANFs, viz., nitriles, thiocyanates, isothiocyanates, and oxazolidinethione (OZT) (Rouzaud, 2004). Food preparation practices such as soaking and cooking can irreversibly destroy the inherent myrosinase activity thereby significantly reducing the amount of ANFs (Samtiya et al., 2020). Despite thermal inactivation of the endogenous myrosinase activity, consumption of cooking-processed GSLs can still lead to the formation of the breakdown products with the supplementary myrosinase activity arising from the gut microbiota (Narbad and Rossiter, 2018). Although the GSL breakdown products have been reported to interfere with thyroid metabolism in farm animals (Fenwick et al., 1983); there are several studies supporting the antitumoral, proapoptotic, and anti-invasive activities of these compounds relevant to human health (Vig et al., 2009). Reportedly, these products induce Phase II detoxification enzymes (quinone reductase, glutathione-S-transferase, and glucuronosyl transferases), and exert an antiproliferative regime against several types of cancer cells (Wu et al., 2009).

In Assam, rapeseed-mustard seeds are fermented through the solid-state spontaneous process under contrasting pH environments (alkaline and acidic) to generate two gastronomically different products viz., *Kahudi*, a pungent sour paste resulting from acidic fermentation, and *Kharoli*, a pungent paste with umami feel as an outcome of alkaline fermentation (Goswami et al., 2017). The resulting products are coarse granular pastes and are consumed as side dishes of the main meals. Fermented oilseed products such as *Kahudi* and *Kharoli* are very unique, and regularly consumed by

most of the communities of Assam. Although indigenous fermentation appears as random and spontaneous, the preparation practices such as soaking, drying, pan-frying, salting, etc. followed by concoction with additives dictate a cascade of biochemical changes in the substrate which influences the fermenting microbiota structure as well as determines the final nutritional and sensory qualities. Very little is known about the microbial diversity structure as well as metabolites present in these fermented oilseed products. In the present study, we integrated 16S ribosomal ribonucleic acid amplicon sequencing (through Illumina-based Next-Gen sequencing) and liquid chromatography-mass spectrometry (LC-MS)-based metabolomics to evaluate the effects of preparation practices (especially, additive types) on overall microbiota and metabolome of these two fermented food products, i.e., *Kahudi* and *Kharoli*. Comprehensive knowledge of the metabolite/phytochemical profile and microbiota associated with these products will not only aid in optimizing the quality but also increase their commercialization prospects.

## Materials and methods

### Traditional preparation of *Kahudi* and *Kharoli*

*Kolakhar* (an alkaline food additive) and *Thekera* (ripe fruits of *Garcinia pedunculata* L.) are the two basic ingredients of *Kharoli* and *Kahudi*, respectively which determine the type of fermentation (acidic or alkaline) of the oil seeds in Assam. *Kharoli* involves alkaline (pH 8.0–8.5) solid-state fermentation owing to the alkaline nature of *Kolakhar*. *Kolakhar* is an alkaline condiment prepared preferably from dry banana peels, banana pseudostem, and rhizome (*Musa paradisiaca*, family: *Musaceae*; vernacular name: *Bheem-Kol*). As a general practice, the ashes of sun-dried mature banana peels are dipped in a cup of water and filtered using a soft muslin cloth. *Khar* has a black-tea-like appearance with colloidal consistency (pH of  $10.24 \pm 0.02$ ). The proximate composition of *Khar* included: total dissolved solids (TDS):  $6,800.0 \pm 25.2$  mg/l; total hardness of  $285.0 \pm 2.5$  mg/l; calcium:  $338.02 \pm 24.12$  mg/l; potassium:  $168.0 \pm 16.82$  mg/l; magnesium:  $146.51 \pm 5.56$  mg/l; manganese:  $193.45 \pm 11.80$  mg/l; zinc:  $12.65 \pm 1.55$  mg/l, and iron:  $0.66 \pm 0.02$  mg/l.

For *Kharoli* preparation, the mustard seeds are first washed properly, drained, and pan-fried, followed by a thorough pounding with the occasional addition of mustard oil during the process. The slow frying step breaks down the chemical bonds in protein-lipid moieties thereby liberating the oil from the lipoprotein molecules. This step also inactivates myrosinase resulting in high yield and less pungent oil. The oily batter is then mixed with final portions of mustard oil and *Khar* with pinches of table salt. The concoction is shaped into small balls and wrapped in banana leaves (preferably leaves



FIGURE 1

Culinary practice of *Kahudi* and *Kharoli* preparation; (A) slow roasting of mustard seeds; (B) the acidic additive: dried mangosteen peels soaked in water; (C) the alkaline additive, *Khar* (D) mixing of additive (mangosteen juice in case of *Kahudi*; *Khar* in case of *Kharoli*) (E) dough shaped into small balls (F) dough-balls wrapped in banana leaves; ready for fermentation.

of *Musa paradisiaca*, made malleable over the fire flame). The wrapped packets are placed inside bamboo cylinders (*Bahor-Chunga*) and allowed to ferment over a fireplace (*Dhuwa-Chang*) for 5–11 days. The preparation of another mustard-seed fermented product named *Kahudi* is very similar to *Kharoli* except that mangosteen (*Thekera-Tenga*) replaces *Khar* in its preparation. The process involves pounding rapeseed with extracts of *Thekera-Tenga* (*Garcinia pedunculata*, ripe fruits that are sour) into a mash (Figure 1). Mangosteen contained a significant amount of ascorbic acid ( $141.0 \pm 6.3$  mg per 100 g of dry weight),  $\beta$ -carotene ( $10.6 \pm 0.8$  mg per 100 g of dry weight), polyphenols ( $18.63 \pm 2.5$  mg per 100 g of dry weight), tannins ( $330.0 \pm 18.2$  mg per 100 g of dry weight), and moisture (89.71%). *Kahudi* and *Kharoli* are pungent and strong in taste and are mostly served as a condiment to an Assamese meal.

### Sampling of traditionally prepared *Kahudi* and *Kharoli*

With the help of a local resource person, the preparation of these fermented products was commissioned to three village rural households located at the Jorhat West revenue circle (Jorhat district), State of Assam, India (see Supplementary Table S1 and Supplementary Figure S1 for the details of the sampling sites). The fermentation period for both *Kahudi* and *Kharoli* shows a seasonal variation (mostly, 5–7 days in summer and 9–11 days in winter). As recommended by the local resource person, the sampling was done on

the 7th day after visual and sensorial inspection of the end product. Triplicate samples of each food product were collected from three individual households. Equivalent weights of these triplicates from a single household were pooled to constitute one composite sample. In this way, a total of six composited naturally fermented mustard-seed product samples (three for each; i.e., *Kahudi* and *Kharoli*) representing the biological replicates for *Kahudi* and *Kharoli* were constituted. These samples were carefully placed on ice packs and aseptically transported to the laboratory for immediate biochemical (proximate analysis and LC-MS-based studies) and microbiological analyses.

### Proximate composition and culture-dependent microbiological analysis

The pH values of the mustard seeds, *Kahudi*, and *Kharoli* (10 g sample), as well as the two additives (i.e., *Khar* and *Mangosteen* solution), were measured with the help of a digital pH meter (model: pH700, Eutech instruments, Singapore). The crude protein, crude fiber, total starch, and soluble protein were determined by following standard methods (Horwitz et al., 2005). The mineral composition (*viz.*, calcium, potassium, magnesium, manganese, zinc, and iron) of the two fermented food sample types were determined using an ICE 3,000 Series atomic absorption spectrophotometer (Thermo Scientific, USA) with known concentrations of experimental standards (Viñas,

1993). For the estimation of tannins, the samples were digested with concentrated sulphuric acid through the micro-Kjeldahl method, and the color development after the addition of the Folin–Denis reagent was recorded in a spectrophotometer (Hitachi 2205) at 760 nm (Katoch, 2011). The amount of phytate was determined by capturing the  $\text{Na}_2\text{SO}_4$ -treated contents (2.0 g sample treated in 100 g/l  $\text{Na}_2\text{SO}_4$  supplemented with 1.2% HCl) in an anion resin column (Cosgrove, 1963). Total viable count (TVC) was enumerated on the Plate Count Agar (Merck, Germany) incubated at 28°C for 48 h. The presence of any coliform bacteria (CFM) and *Escherichia coli* (EC) was evaluated on Chromogenic Chromocult coliform agar (Merck, Germany) incubated at 37°C for 24 h. All the experiments were performed in triplicate; raw mustard seeds were included in the proximate analysis to assess fermentation-induced compositional changes in the seeds. The variations (analysis of variance) in the data were evaluated using the XLSTAT package program. The differences in average were evaluated by the least significant difference test.

## Liquid chromatography-mass spectrometry (LC-MS) aided characterization of major metabolites

The extract of *Kahudi* and *Kharoli* (each dissolved separately in methanol) –5 mg/ml was filter sterilized. The LC-ESI-MS was carried out in the Agilent 1,260 Infinity HPLC system (Agilent Technologies, USA) equipped with an autosampler by loading each sample with an injection volume of 20  $\mu\text{l}$  into the system. The ESI positive was obtained in an ODS-2 column (250  $\times$  4.5 mm, pore size: 5  $\mu\text{m}$ ) with a gradient elution condition consisting of mobile phase: (a) acetonitrile (solvent A) (b) 0.01% formic acid in water (gradient mode, solvent B). The gradient profile was as follows: 95% B at 0–5 min, 70% B at 6–11 min, 40% B at 12–19 min, continued with 20% B till 25 min, 95% B continued till 30 min. The flow rate was 1.5 ml/min. The scan range was from 80 to 1,000 m/z. The major metabolites/compounds were identified by calculating the molecular mass against a particular retention time. The peaks from the ESI positive containing the m/z ratio in one of the axes and the relative abundance (%) in the other axis for different retention times were used to predict the compounds. The molecular mass for each retention time was calculated by identifying the different adducts like M+H, M+Na, M+CH<sub>3</sub>OH +H, M+ACN+Na, M+ACN+H, M+2Na–H, 2M+K, M+2K+H, 2M+Na, and 2M+NH<sub>4</sub>. The molecular masses thus obtained were compared with the molecular weight of the reference compounds from different sources including the PubChem, NIST molecular database, and other published work (Nicolescu, 2017).

## Univariate and multivariate analyses of LC-MS data

The Wilcoxon tests and paired *t*-tests were performed to compare relative abundances of the phytochemicals/metabolites present in *Kahudi* and *Kharoli*. As per Bonferroni's adjustment, the *p*-value for the multiple tests involving 26 parameters (metabolites) was corrected and considered to be significant when *p* < 0.0019. The Multivariate Data Analysis Software SIMCA<sup>®</sup> 17 was used to analyze metabolomics data. The background noise and high variances in the variables were minimized by the logarithmic transformation of the data file. The similarities or dissimilarities in the sample profile were assessed by performing an unsupervised principal components analysis (PCA). A two-dimensional score plot was plotted containing the principal components which represented the correlated spectral differences. The PCA model's goodness of fit was quantitatively measured by the parameters R2X (goodness of fit) and the predicted variation Q2X (goodness of prediction). The contribution plot of scores was created to identify the major variables (metabolites) which contributed significantly to the PCA model. In addition to principal component analysis which is naturally unsupervised, we also performed supervised Orthogonal Partial Least-Squares Discriminant Analysis (OPLS-DA). This model technique can effectively take care of large numbers of variables, small sample sizes, and multicollinearity (Trygg and Wold, 2002). The visualization of the OPLS-DA model was achieved through the generation of a score plot (iterations at 200) which is useful in the visual inspection of grouping and trends in the data.

## DNA extraction and barcoding pcrs for MiSeq sequencing

For metagenomic DNA extraction, 10.0 g of the samples were blended for 30 mins in a 40 ml sterile water solution supplemented with 1% Tween 80. The turbid supernatant (35 ml) was collected in a fresh tube and the suspended bacterial cells were harvested in the form of a pellet through rapid centrifugation at 3,000  $\times$  g for 5 min at 4°C. Genomic DNA samples from these pellets were extracted following the manufacturer's protocol mentioned in the Environmental gDNA isolation kit (Xcelgen, India) (Ghosh and Das, 2018). The amplicon libraries were constructed following two rounds of PCR amplification. The first amplification of the ~450-bp V3–V4 hypervariable regions of the bacterial 16S rRNA gene was performed with the primers V3-F (5'-ACGGRAGGCWGCAGT-3') and V4-R (5'-TACCAGGGTATCTAATCCT-3') (Klindworth et al., 2013). The conditions for the first round of PCR were as follows: initial denaturation at 94°C for 1 min, followed by 30 cycles of amplification at 94, 65, and 72°C each for 1 min, and a

final extension step at 72°C for 10 min. This round of PCR was performed in triplicate, and the amplified products were pooled and purified with the AmPure XP beads. An aliquot of 100 ng purified DNA was considered for the second round of PCR (in triplicate), where sample multiplexing was sieved through the use of the same primers i.e., V3-F and V4-R supplemented with an overhang adapter. The number of PCR cycles was limited to 12 for the second round of PCR with other temperature/duration conditions remaining the same as the first round. The final PCR products were again purified with the AmPure XP beads and the integrity of the PCR products was checked on a DNA-1000 chip on the Agilent Bioanalyzer 2100 system. In the next step, the purified amplicons were indexed through Indexing PCR with Nextera XT Indices as Index 1. After quality checking of the resulting Index PCR products on the DNA-1000 chip on the Agilent Bioanalyzer 2100, and quantification by using the Qubit HS and qPCR, the libraries (600 µl) were finally loaded on the Illumina MiSeq cartridge for cluster generation by hybridization. The manufacturer's instructions mentioned in the complementary Illumina Reagent Kit were followed for this purpose. The immobilized amplicons (hybridized onto the oligonucleotide-coated surface of the flow cell) served as templates for the generation of clonal DNA clusters (2 × 250 bp; PE). The FastQ data files generated at the end of the run (MiSeq Reporter software, Illumina, USA) were collected for downstream bioinformatics analysis. With the help of the multiplexing indices, the paired-end sequences were demultiplexed to their respective sample of origin. Raw read sequences (a total of six pairs) were deposited to the Sequence Read Archive under the Bio project id PRJNA742190.

### Quality filtering, the ASVs, and taxonomic assignment

The quality of the sequence reads was first examined using the FastQC followed by importing the demultiplexed paired-end fastq data as QIIME 2 artifacts into the QIIME 2 2018.11 environments. The sequence reads were quality-filtered and denoised (removed reads without primer sequences) using the DADA2 pipeline through the q2-dada2 plugin (Callahan et al., 2016). This step also trimmed the Illumina adaptor and barcode sequences from the reads. Next, both forward and reverse reads were merged into a single sequence followed by quality filtering to remove low-quality ends (threshold set at 200 base pairs). The joined sequences were denoised to eliminate low-quality bases and to generate amplicon sequence variants (ASVs) using the SILVA rRNA database (release 132) (Quast et al., 2013). The taxonomic assignment was performed using the “classify-sklearn” function, a pre-fitted sklearn-based taxonomy classifier.

### Statistical analysis of amplicon sequencing data

The QIIME2 script “multiple\_rarefactions.py” was used to rarefy the ASVs at a depth of 50–6,482 sequences and to create alpha diversity rarefaction curves. The “make\_rarefaction\_plots.py” script which is based on the Kruskal-Wallis test was used to generate alpha diversity indices (observed features, Shannon index, and Faith's Phylogenetic Diversity index). Individual PCoA biplots were created based on beta diversity measures (Bray-Curtis, Jaccard, Unweighted- and Weighted UniFrac distance matrices) and visualized through the Emperor tool. A Permutational Multivariate Analysis of Variance (PERMANOVA) test (based on the Bray-Curtis distance) was used to compare the bacterial communities between the two food samples. A significant PERMANOVA test further led to assessing the heterogeneity of multivariate dispersion through a Permutational Analysis of Multivariate Dispersion (PERMDISP) test. The variation in the bacterial community and metabolite abundance was evaluated by principal component analysis (PCA) using XLSTAT software.

## Results

### Biochemical and nutritional profile of *Kahudi* and *Kharoli*

The data on physicochemical parameters (pH, soluble protein, crude fiber, total starch, and anti-nutritional factors as well as micronutrient status) of the two fermented samples, and their raw substrate i.e., mustard seeds are presented in Table 1. Owing to the extreme pH values of the additives, i.e., *Khar* solution and *Garcinia* juice (i.e.,  $9.81 \pm 0.02$  and  $4.3 \pm 0.02$ , respectively), the pH values of the two fermented food products i.e., *Kharoli* and *Kahudi* were also in mild alkaline (pH:  $7.6 \pm 0.03$ ) and mildly acidic ( $5.92 \pm 0.02$ ) ranges, respectively. Fermentation significantly increased the contents of crude protein (36.7 and 44.9% higher crude proteins in *Kahudi* and *Kharoli*, respectively as compared to the raw unfermented substrate), while decreasing the amounts of crude fiber, total carbohydrate, and anti-nutritional factors (such as tannins and phytates). The level of calcium was  $357.82 \pm 2.4$  mg/100 g in *Kahudi* and  $472.34 \pm 5.11$  mg/100 g in *Kharoli*; these values in the two products were 42.4% and 88.0% higher (statistically significant,  $p < 0.05$ ) than the raw substrate ( $251.2 \pm 4.8$  mg/100 g). The levels of potassium were also statistically higher in the food products (*Kahudi*:  $818.81 \pm 13.9$  mg/100 g and *Kharoli*:  $1,132.86 \pm 14.5$  mg/100 g) than in the substrate ( $721.5 \pm 11.4$  mg/100 g). Apart from the aforesaid minerals, the fermented samples also harbored a considerable amount of manganese, zinc, and iron (range for manganese:  $17.94 \pm 1.21$  to  $16.93 \pm 1.4$  mg/100 g; range for zinc:  $12.01 \pm 0.47$  to  $21.55 \pm 1.8$  mg/100 g; range for iron:  $67.07 \pm 2.9$  to  $86.21 \pm 2.4$  mg/100 g).

TABLE 1 Proximate composition, micro- and macronutrient status of *Kahudi* and *Kharoli* (with original substrate for fermentation, mustard seeds).

Item	pH	Crude protein (g/100g)	Crude fiber (g/100g)	Total carbohydrate (g/100g)	Tannins (mg/100g)	Phytate (mg/100g)	Minerals (mg/100g)				Total viable counts [log(CFU/g)]	<i>Escherichia/Fecal coliforms</i>	
							Calcium (Ca)	Potassium (K)	Manganese (Mn)	Zinc (Zn)			Iron (Fe)
Mustard seeds		26.7 ± 1.8 <sup>c</sup>	10.77 ± 0.74 <sup>a</sup>	28.3 ± 1.5 <sup>a</sup>	1.65 ± 0.11 <sup>a</sup>	42.47 ± 2.2 <sup>a</sup>	251.2 ± 4.8 <sup>c</sup>	721.5 ± 11.4 <sup>c</sup>	2.5 ± 0.24 <sup>c</sup>	6.15 ± 0.84 <sup>c</sup>	9.05 ± 0.45 <sup>c</sup>	0.11 ± 0.01 <sup>c</sup>	ND
<i>Kahudi</i>	5.92	36.5 ± 2.2 <sup>b</sup>	4.45 ± 0.24 <sup>c</sup>	6.5 ± 0.77 <sup>b</sup>	0.56 ± 0.08 <sup>b</sup>	1.4 ± 0.02 <sup>b</sup>	357.82 ± 2.4 <sup>b</sup>	818.81 ± 13.9 <sup>b</sup>	17.94 ± 1.2 <sup>a</sup>	12.01 ± 0.47 <sup>b</sup>	67.07 ± 2.9 <sup>b</sup>	5.04 ± 0.74 <sup>b</sup>	ND
<i>Kharoli</i>	7.6	38.7 ± 2.1 <sup>a</sup>	4.91 ± 0.31 <sup>b</sup>	6.1 ± 0.78 <sup>c</sup>	0.34 ± 0.05 <sup>c</sup>	1.2 ± 0.01 <sup>c</sup>	472.34 ± 5.1 <sup>a</sup>	1132.86 ± 14.5 <sup>a</sup>	16.93 ± 1.4 <sup>b</sup>	21.55 ± 1.8 <sup>a</sup>	86.21 ± 2.4 <sup>a</sup>	5.80 ± 0.19 <sup>a</sup>	ND

N.B. For all variables sharing the same superscript, the difference between the means is not statistically significant.

## LC-MS-based non-targeted metabolomics showed a distinct metabolite abundance profile

A total of 24 compounds were identified to be present in both *Kahudi* and *Kharoli*. 12 compounds could be assigned to three groups; viz., (i) glucosinolates: 4 compounds; (ii) glucosinolate breakdown products: 6 compounds, and (iii) other phytochemicals non-linked to glucosinolates: 14 compounds. A few of the therapeutically and nutritionally important compounds were sinapine (MW-310.37), indole-3-carbinol (MW 147.18), anthocyanin (MW 207.24), oxazolidinethione (MW 103.15), erucic acid (MW 338.57), sulforaphane (MW 177.29), 2-allyl- glucosinolates (MW 359.4), allyl isothiocyanate (MW 99.15), phytate (MW 682.02), ascorbic acid (MW 176.2) and  $\beta$ -carotene (MW 536.87). These important compounds were marked in the chromatogram against their respective retention time (Supplementary Figure S2). For the compounds isolated and identified in this paper, their prominent diagnostic mass spectral ions, relative intensities, and variable importance on projection (VIP) are presented in Table 2.

## Additive type and pH influenced the fermentation fate of processed mustard-seeds

Multivariate statistical analysis was performed to elucidate the similarity and dissimilarity in the metabolite profiles of *Kahudi* and *Kharoli*. In PCA analysis, a total of six principal components were obtained; the variables (metabolites) with similar abundance profiles were proximately projected in the PCA space. The metabolites linked to *Kahudi* were located on the right side of the confidence interval, while the *Kharoli* metabolites were placed on the left side of the confidence interval (Figure 2A). Abundance levels of oxazolidinethione and 2-allyl-glucosinolate were similar in both the products for which these two metabolites were placed at the middle of the confidence interval. The contribution score plot (Figure 2B) illustrates the relative contribution of each variable to the PCA model. High abundance of a few metabolites viz., sinapine, indole-3-carbinol,  $\gamma$ -linolenic acid, etc. in *Kahudi* and metabolites viz.,  $\beta$ -sitosterol acetate, 3-butylene glucosinolate, erucic acid, glycine, etc. in *Kharoli* significantly contributed to the PCA plot. However, the cumulative R2X = 0.539, Q2 = 0.454, (i.e., cumulative PC1 and PC2 coverage of 53.9%; Q2 < 0.5) indicated that the model was not very effective in explaining all the variations in the original state of metabolomic data. In such a situation, the PLS/OPLS-DA modeling becomes handy as this supervised approach takes care of orthogonal variables which are not linked to any categorical data in the metabolite profiles and therefore, as

TABLE 2 List of few major metabolites identified in *Kahudi* and *Kharoli*.

Compound	Retention time	Predicted molecular mass	Relative intensity in <i>Kahudi</i> (in %)	Relative intensity in <i>Kharoli</i> (in %)	VIP score
<b>Major glucosinolates</b>					
3-butylene-3-glucosinolates	21.94	372.4	1.45 ± 0.76 <sup>b</sup>	17.33 ± 1.53 <sup>a</sup>	-1.12451
3-butylene glucosinolates	13.49	373.4	2.34 ± 0.56 <sup>b</sup>	20.33 ± 1.53 <sup>a</sup>	-1.46565
2-allyl-glucosinolates	22.86	359.4	26.33 ± 1.53 <sup>b</sup>	32.0 ± 2.0 <sup>a</sup>	0.584382
2-hydroxy-3-butylene glucosinolates	23.99	389.4	1.14 ± 0.21 <sup>b</sup>	22.67 ± 1.53 <sup>a</sup>	-1.4088
<b>Glucosinolates breakdown products</b>					
Indole-3-carbinol	10.83	147.18	36.67 ± 1.53 <sup>a</sup>	6.0 ± 1.0 <sup>b</sup>	2.7364
Allyl isothiocyanate	11.07	99.15	17.0 ± 2.0 <sup>a</sup>	5.67 ± 1.15 <sup>b</sup>	0.958873
Erucic acid	15.33	338.57	8.0 ± 1.0 <sup>b</sup>	40.0 ± 2.0 <sup>a</sup>	-1.86342
Sulforaphane	21.3	177.29	13.33 ± 0.58 <sup>a</sup>	9.00 ± 1.00 <sup>b</sup>	0.424619
Oxazolidine thione	11.82	103.15	8.0 ± 1.0 <sup>b</sup>	11.33 ± 1.53 <sup>a</sup>	-0.14531
Sinapine	10.42	310.37	28.67 ± 3.06 <sup>a</sup>	12.0 ± 2.0 <sup>b</sup>	1.78565
<b>Other important phytochemicals</b>					
Ascorbic acid	9.91	176.2	6.0 ± 1.0 <sup>a</sup>	1.87 ± 0.45 <sup>b</sup>	0.298277
Campesterol	9.94	400.37	1.75 ± 0.15 <sup>b</sup>	8.0 ± 2.0 <sup>a</sup>	-0.78337
Anthocyanin	11.47	207.24	9.0 ± 1.0 <sup>b</sup>	18.0 ± 1.0 <sup>a</sup>	-0.38953
$\beta$ -Sitosterol acetate	13.39	456.1	3.66 ± 0.28 <sup>b</sup>	21.67 ± 2.08 <sup>a</sup>	-1.57937
Phytate	13.59	682.02	11.33 ± 1.53 <sup>a</sup>	1.33 ± 0.61 <sup>b</sup>	0.758884
B-carotene	13.9	536.87	11.33 ± 0.58 <sup>a</sup>	1.08 ± 0.18 <sup>b</sup>	0.807347
Oleic acid	14.03	304.24	1.25 ± 0.30 <sup>b</sup>	10.33 ± 1.53 <sup>a</sup>	-0.89708
Ononitol	14.17	554.28	14.0 ± 1.0 <sup>a</sup>	1.22 ± 0.78 <sup>b</sup>	1.13057
Clionasterol acetate	14.58	456.4	2.10 ± 0.03 <sup>b</sup>	11.0 ± 2.0 <sup>a</sup>	-0.84023
Sitosterol	15.12	414.39	20.67 ± 1.15 <sup>b</sup>	37.0 ± 1.0 <sup>a</sup>	-0.40321
$\gamma$ -Linolenic acid	20.75	331.29	29.0 ± 1.0 <sup>a</sup>	2.14 ± 0.39 <sup>b</sup>	2.5244
Glycine	22.05	291.15	1.44 ± 0.25 <sup>b</sup>	25.33 ± 1.53 <sup>a</sup>	-1.74994
O-pentafluoropropionyl-dihydrocapsaicin	23.34	453.19	6.33 ± 0.58 <sup>a</sup>	0.98 ± 0.05 <sup>b</sup>	0.342736
Clerosterol acetate	23.51	454.38	5.67 ± 0.58 <sup>a</sup>	1.48 ± 0.24 <sup>b</sup>	0.298277

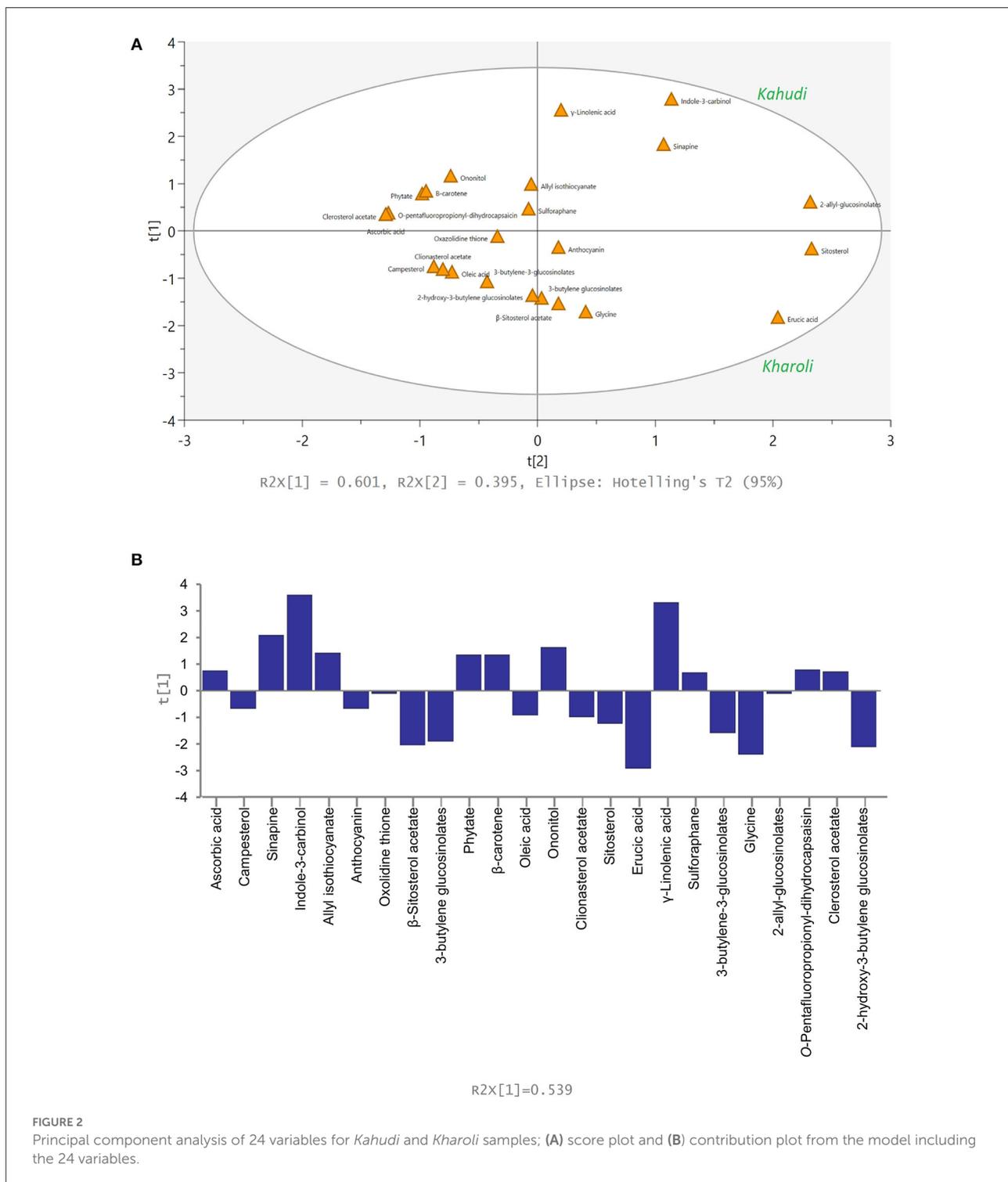
N.B. For all variables sharing the same superscript, the difference between the means is not statistically significant.

compared to the PCA, becomes more inclusive in explaining the underlying variations between the two products. As shown in the OPLS-DA model (Figure 3), most of these metabolites were projected within the 95% confidence interval. Pareto scaling of the OPLS-DA model explained 98.0% [i.e. (R<sup>2</sup>Y(cum)) = 0.980] of the variations in the metabolite abundance profiles. The predictive value of the model was found to be close to 1.0 [Q<sup>2</sup>(cum) = 0.969] indicating that the model was effective in projecting the variables in PC space with no overfitting phenomenon. Analysis of variance (CV-) ANOVA showed a low *p*-value of 0.0022 (lesser than the *p*-cutoff value <0.05) confirming the cross-validation performance of the model. The contribution plot of the model identified the relative importance of each metabolite in differentiating the two fermented food products. Both OPLS-DA and PCA detected the same metabolites to be discriminative. Overall, it can be

concluded that additive type and pH influenced the fate of mustard-seed fermentation.

## Overall bacterial community structure

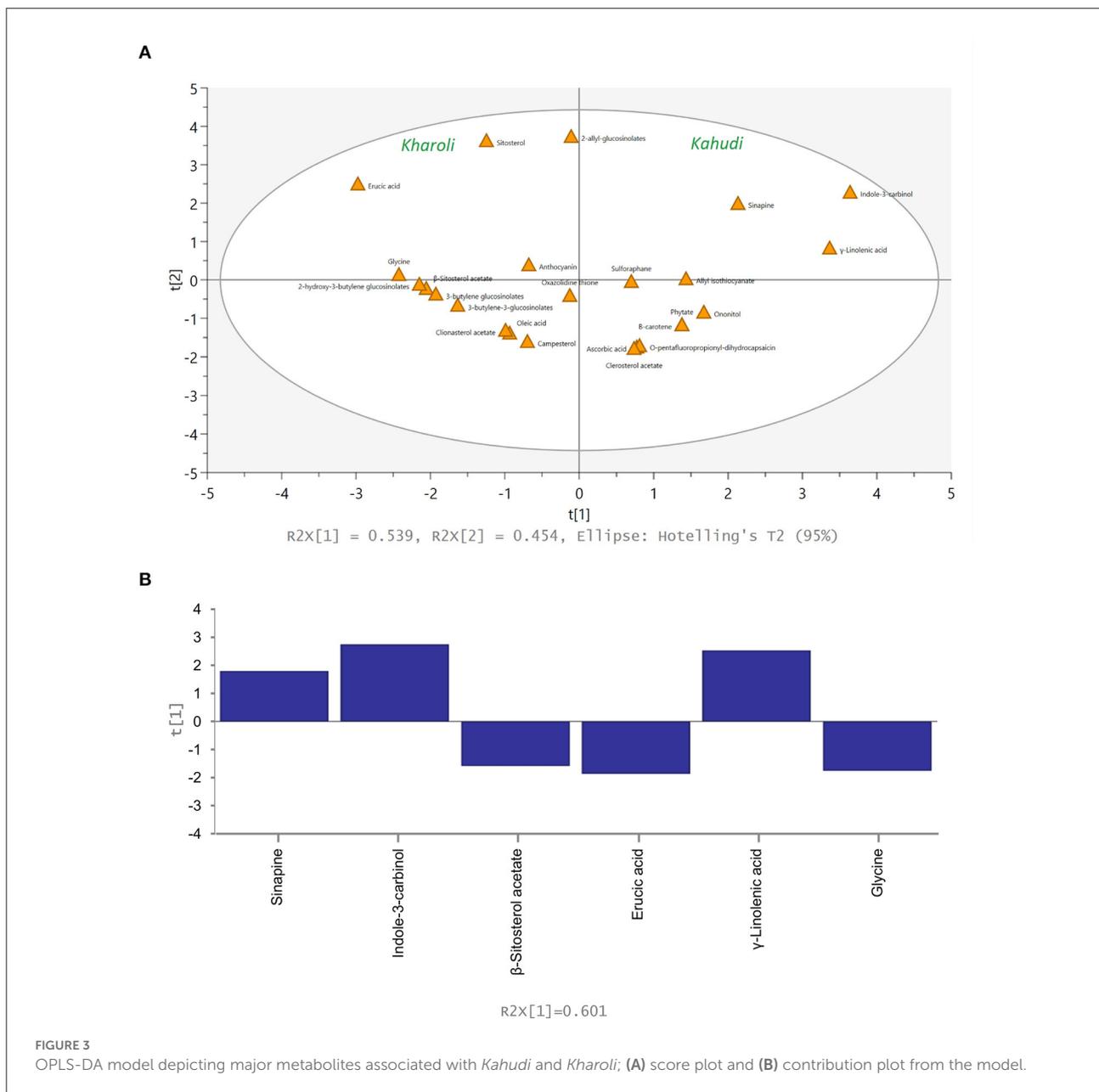
The 16S rRNA (V3–V4) amplicon sequencing approach was employed to assess the abundance of bacterial taxa in the two traditional fermented foods. After preliminary filtering and adaptor trimming, samples from *Kahudi* accounted for a total of 678,778 paired-end (PE) reads with a median of 168,593 PE reads per sample; while *Kharoli* had a total of 705,561 PE reads with a median of 170,349 PE-reads. After preprocessing, i.e., removal of low-quality PE-reads, chimeric and non-linked sequences, 676,612 clean PE-reads from *Kahudi* and 700,753 clean PE-reads from *Kharoli* were classified into 209 and 196 ASVs respectively.



### Multivariate analysis of bacterial flora showed stringent fermentation environment

The plateau-reaching rarefaction curves of the amplicon sequence variants (ASVs) and Shannon index

(Supplementary Figure S3) indicated that the sequencing approach was successful in capturing the majority of the bacterial diversity in the fermented food samples. Alpha diversity indices (observed features, Shannon, Faith's phylogenetic diversity) were compared between the two food-type samples through the Kruskal-Wallis statistical



test. Both *Kahudi* and *Kharoli* harbored equivalent species richness (*Kahudi*:  $10.67 \pm 2.49$ ; *Kharoli*:  $10.67 \pm 3.79$ ) (Supplementary Table S2). However, the average Shannon diversity index for *Kharoli* was more than that of *Kahudi* (*Kharoli*:  $0.94 \pm 0.00$ ; *Kahudi*:  $0.36 \pm 0.01$ ;  $p$ -value: 0.049; difference statistically significant). This indicated that a small number of ASVs (most assigned to *Bacillus* species) in *Kahudi* dominated the microbiome structure resulting in lesser evenness in the samples compared to *Kharoli*. The Principal coordinate analysis (PCoA) was performed to evaluate the differences in genus-level taxonomic groups using the Bray-Curtis, Jaccard, Unweighted- and Weighted UniFrac distance

matrices (Supplementary Figure S4). The *Kharoli* samples clustered to one end of principal component (PC) 1 in both the Jaccard and Unweighted UniFrac distance plots, while all the *Kahudi* samples were placed to the opposite ends (PC1 + PC2: 74.72 and 87.14% of variation explained for Jaccard and Unweighted UniFrac distances, respectively). We observed that the microbial community structure significantly differed between the two food samples (PERMANOVA,  $P = 0.095$ ) (Supplementary Figure S5). A non-significant PERMDISP test ( $p = 0.571$ ) indicated homogeneity of dispersion and the difference in beta diversity to rely on the additive-induced pH changes (Supplementary Figure S6). It is possible that

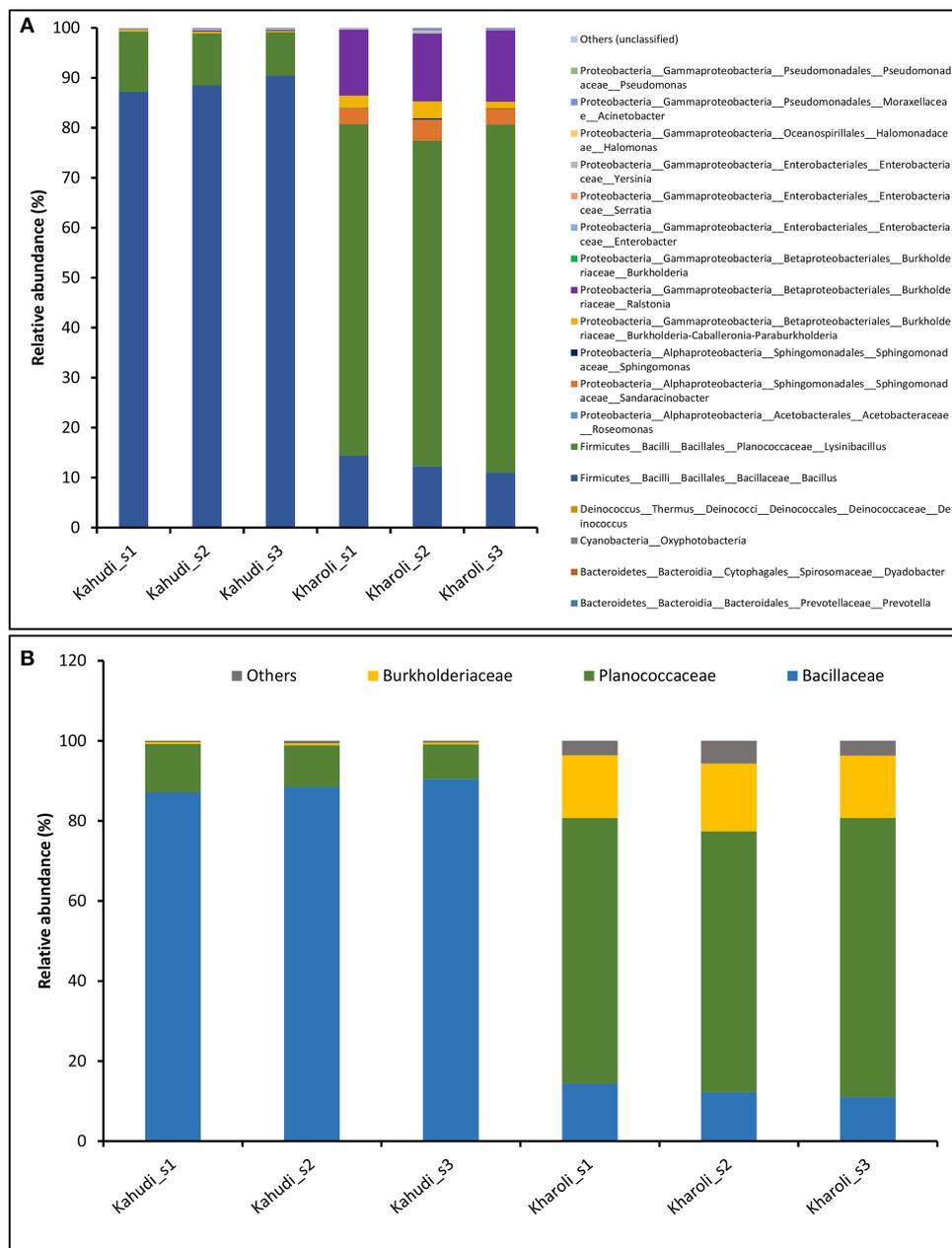
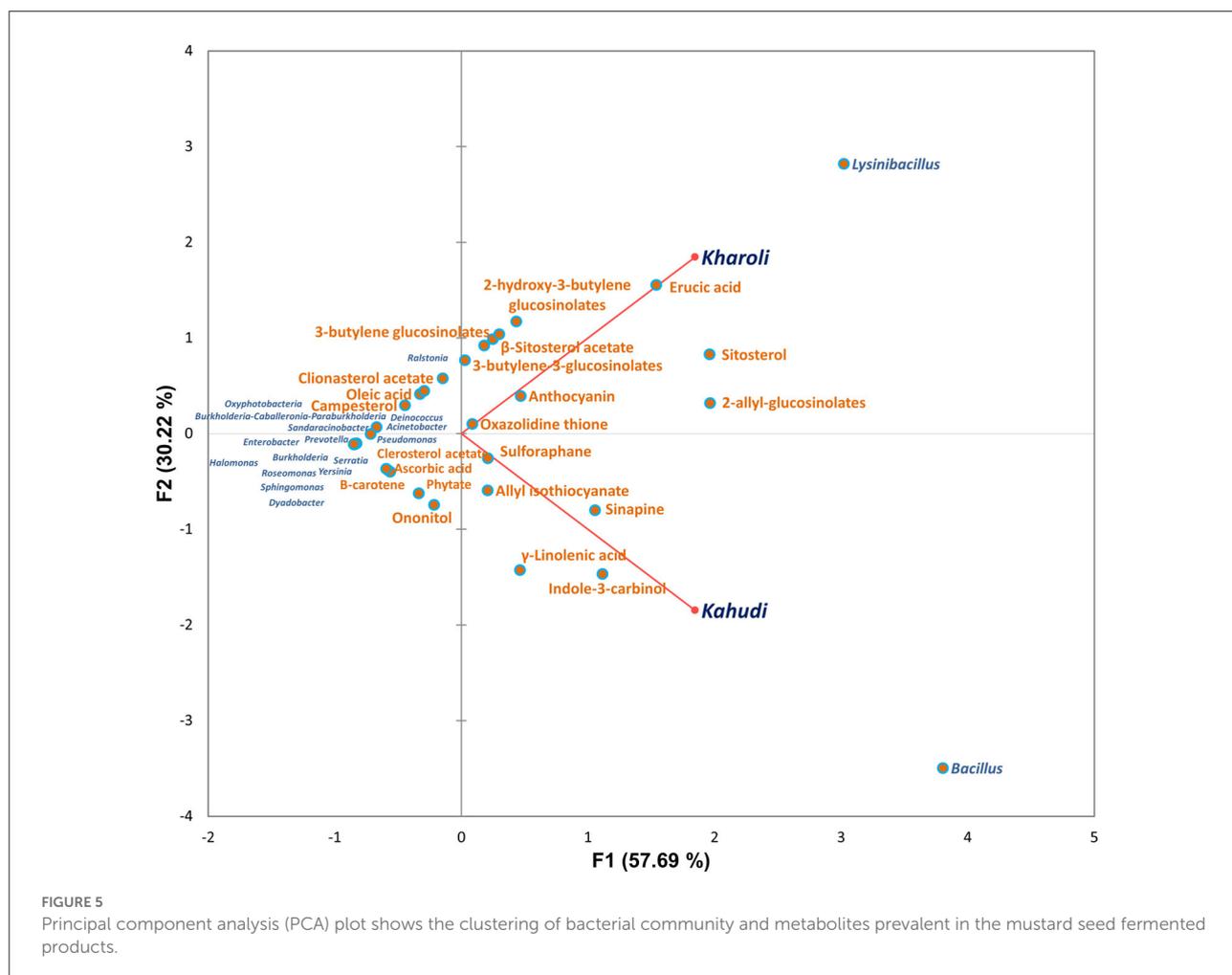


FIGURE 4 The overall bacterial composition of *Kahudi* and *Kharoli* at different taxonomic levels; (A) genus and (B) family.

additive-induced pH modulated the microbial abundance profiles in these food samples. The PCoA analysis also revealed significant differences (ANOSIM,  $p = 0.005$ ,  $R = 0.11$ ) in bacterial diversity between the two sample types; however, there was no significant difference among the same type of products (Supplementary Figure S7). This indicated that despite locational differences in sampling, the preprocessing before fermentation renders a stringent environment for bacterial microbiota to survive and flourish in mustard seeds.

### Potential stress-resilient bacteria colonized and dominated the mustard seed fermentation microbiota

The bacterial population in *Kahudi* was majorly dominated by the phylum Firmicutes ( $99.1 \pm 0.18\%$ ) with a minor population of Proteobacteria ( $0.90 \pm 0.19\%$ ). In the case of *Kharoli*, the major bacterial taxa were distributed between Firmicutes ( $79.6 \pm 1.92\%$ ) and Proteobacteria ( $20.37 \pm 1.94\%$ ).



Members of the phylum *Deinococcus-Thermus* also had low abundance (<1% in both) in both the fermented foods samples. Phylum *Firmicutes* was represented by four families belonging to *Bacillaceae* ( $88.7 \pm 1.67\%$  in *Kahudi*;  $12.51 \pm 1.75\%$  in *Kharoli*) and *Planococcaceae* ( $10.4 \pm 1.74\%$  in *Kahudi*;  $67.1 \pm 2.37\%$  in *Kharoli*). Phylum *Proteobacteria* included *Burkholderiaceae* only ( $0.47 \pm 0.04\%$  in *Kahudi*;  $16.09 \pm 0.75\%$  in *Kharoli*); *Sphingomonadaceae* ( $3.65 \pm 0.75\%$  in *Kharoli*; less than  $0.02\%$  in *Kahudi*); *Moraxellaceae*, *Enterobacteriaceae*, *Pseudomonadaceae*, *Halomonadaceae*, *Acetobacteraceae*. At the genus level, *Bacillus* ( $88.7 \pm 1.67\%$ ) followed by *Lysinibacillus* ( $10.4 \pm 1.74\%$ ) dominated the microbiota in *Kahudi*; while, in the case of *Kharoli*, it was vice versa (*Bacillus*:  $12.5 \pm 1.75\%$ ; *Lysinibacillus*:  $67.1 \pm 2.37\%$ ) (Figure 4).

## Relationship between microbial community structure and prevalent metabolites

A combinatorial Principal component analysis (PCA) involving microbiota structure at the genus level and metabolite

abundance showed that each fermented food was unique in terms of the microbial taxa as well as the prevalent metabolites. The plot showed that the *Kharoli* samples were rich in *Lysinibacillus*,  $\beta$ -sitosterol acetate, Erucic acid, sitosterol, 2-allyl-glucosinolates, etc.; while *Kharoli* samples were abundant in *Bacillus*, sinapine, indole-3-carbinol,  $\gamma$ -linolenic acid, etc. in *Kahudi*. This is indicative of the key influence of these microbial taxa on the metabolite abundances, enzymatic functions, and overall fermentation processes in processed rapeseed (Figure 5).

## Discussion

Spontaneously fermented food products prepared from plant-based substrates through alkaline or acidic fermentation are very limited in number and are also lesser known in the world. A handful of such foods reported from Asian and African nations mostly employ soybeans (*Glycine max* L.) and seeds from cultivated as well as wild trees as fermentation substrates (Oguntoyinbo et al., 2016). Various soybean-fermented products consumed in these regions are *Kinema* (consumed by ethnic communities of Himalayan regions), *Thua-nao* (Thailand),

*Douchi* (China), *Soy-dawadawa* (Ghana and Nigeria), etc. (Tamang and Nikkuni, 1996; Dakwa et al., 2005; Wang et al., 2007; Chukeatirote, 2015). Examples of alkaline fermented foods produced from plant seeds are *Dawadawa/Soumbala* (African locust bean), *Dawadawa*-type product (Bambara groundnut; *Vigna subterranea*), *Ogiri Owoh* (Melon, *Citrullus vulgaris* L. and Cottonseed, *Gossypium hirsutum* L.), *Owoh* (African yam bean, *Sphenostylis stenocarpa*), *Ogiri* (Castor oil bean, *Ricinus communis*), *Bikalga* (Roselle; *Hibiscus sabdariffa*), *Ugba* (African oil bean seeds, *Pentaclethra macrophylla*), *Aisa* (rain tree or saman tree, *Albizia saman*), *Okpehe/kpaye/okpiye* (African mesquite, *Prosopis africana*), etc. (Nurudeen and Chimezie Okorie, 2019). Members of the genera *Bacillus*, such as *Bacillus amyloliquefaciens*, *B. brevis*, *B. cereus*, *B. circulans*, *B. firmus*, *B. licheniformis*, *B. pumilus*, *B. licheniformis*, *B. megaterium*, *B. polymyxa*, *B. pumilus*, *B. sphaericus*, *B. subtilis*, *B. thuringiensis* etc. have been reported to be the major dominant taxa of these food products (Ugwuanyi and Okpara, 2020). In this present study, *Bacillus* was found to dominate the microbiota of *Kahudi*. Most of the *Bacillus* species are known to exhibit proteolytic activity which liberates amino acids, essential fatty acids, vitamins, and various other non-protein and soluble nitrogen compounds from the fermenting mass and imparts a characteristic organoleptic and nutritional profile to the food items. The dominant bacterial genus in *Kharoli* was *Lysinibacillus* rather than *Bacillus*. To date, several studies have reported on the occurrence of *Lysinibacillus* in alkaline fermented foods such as *Kinema*, *Bikalga*, *Soumbala*, and also in *Ntoba Mbodi* (a Congolese food made from the alkaline fermentation of cassava leaves) (Sanni et al., 2000; Ouoba et al., 2015). A novel isolate, *Lysinibacillus louembei* sp. nov. NM73T (a designated type strain = DSM 25583T = LMG 26837T) has recently been reported from *Ntoba Mbodi*. The microbiota of *Ntoba Mbodi* was primarily dominated by *Bacillus* spp. (72.2%), especially by various species of the *Bacillus pumilus* group (Ouoba et al., 2015). Important physiological traits such as mesophilic nature, spore-formation, tolerance to environmental stresses (saline-alkaline and saline-acidity), proteolytic and lipolytic ability, etc. enable the survivability and dominance of *Bacillus* and *Lysinibacillus* over other bacterial taxa in solid-state food fermentation (Razzaq et al., 2019). Perhaps, the bacterial species *Bacillus subtilis* draws major attention due to its ubiquitous presence among all these items and also due to its industrial importance as a starter culture for the production of Japanese soybean alkaline fermented delicacy, *Natto* (Sun et al., 2010). In addition, *Lysinibacillus* species (e.g., *Lysinibacillus fusiformis* SCO<sub>2</sub>) also produce another important enzyme known as urethanase which is instrumental in the detoxification of ethyl carbamate (EC) (Liu et al., 2016). This carcinogenic and genotoxic compound is spontaneously formed in food and beverage fermentation. Both *Bacillus* and *Lysinibacillus* spp. are ubiquitous and have been reported from diverse environmental matrixes

(Wenzler et al., 2015). Multiple studies have demonstrated the ability of these species to produce an array of useful enzymes such as alkalohalophilic proteases, thermostable lipases, amylases, etc. (Kalwasińska et al., 2018). Members belonging to the *Bacillus* and *Lysinibacillus* genera can produce spores under unfavorable conditions; the spores are resistant to physical stresses such as wet and dry heat. It is assumed that the resilience of *Bacillus* and *Lysinibacillus* due to their ability to produce stress-tolerant enzymes, spores as well as associations with mustard seeds as endophytic inhabitants enabled them to dominate over other species. Apart from the dominant bacteria, other minor taxa such as *Pseudomonas*, *Acinetobacter*, *Burkholderia*, *Deinococcus*, *Dyadobacter*, *Prevotella*, *Ralstonia*, *Serratia*, *Sphingomonas*, *Roseomonas*, and *Yersinia* were also detected in both the fermented products. These contaminants were probably seed-borne (as seed endophytes) or originated from handling and processing processes.

Fermentation changed the proximate and nutritional composition of raw mustard seeds. The results of the present study showed that the amount of protein was increased with a decrease in total carbohydrate, crude fiber, and anti-nutritional factors (tannins and phytate). The modulation in proximate composition could be attributed to microbial metabolism which metabolically shrinks the carbohydrate source (microbial respiration-induced generation of CO<sub>2</sub>) and anabolically stabilizes the nitrogen source (toward mycoprotein production) which ultimately causes alterations in the C:N ratio. Earlier studies have reported that a mixed culture of *B. subtilis*, *Candida utilis*, and *Enterococcus faecalis* reduced the amounts of crude fiber and phytic acid in rapeseed meals by 25.5 and 42.4%, respectively (Vig and Walia, 2001). In addition to the modulations in proximate composition, there was an elevation in the micronutrient status of the fermented foods. It can be assumed that the addition of the mineral-rich additives i.e. *Khar* and *Garcinia* juice as the main additives as well as extra mustard oil while shaping into dough balls incrementally raised the level of mineral contents in the food items. Additionally, the reduction in anti-nutritional factors such as phytates and tannins also increased the bioavailability of these minerals. LC-MS-based metabolomics study revealed the *Kharoli* samples to contain higher amounts of the major glucosinolates and derivatives viz., 3-butylene-3-glucosinolates, 3-butylene glucosinolates, 2-allyl-glucosinolates (sinigrin), and 2-hydroxy-3-butylene glucosinolates. According to the multivariate analysis, the most relevant metabolites associated with *Kahudi* were high concentrations of sinapine, indole-3-carbinol,  $\gamma$ -linolenic acid, etc.; while, in *Kharoli*, the metabolites were  $\beta$ -sitosterol acetate, 3-butylene glucosinolate, erucic acid, glycine, etc. The term 'glucosinolates' is used for both the glucosinolates and glucoside compounds; hydrolysis of such compounds by the action of endogenous thioglucosidase yields in some bioactive products viz., oxazolidinethione (OZT), thiocyanate, isothiocyanate (ITC), nitriles, and a few other minor derivative

compounds (Holst and Williamson, 2004). *Kahudi* samples had higher amounts of the glucosinolate breakdown products such as indole-3-carbinol, allyl isothiocyanate, sulforaphane, and oxazolidinethione (except for erucic acid which was higher in *Kharoli*). At this point, it was observed that nitriles and thiocyanates could not be detected in any of the samples, mostly because of their thermal instability nature (Hanschen et al., 2014). The absence of thiocyanate instead of isothiocyanates could be attributed to the soaking-activated myrosinase activity and a neutral pH which help in the conversion of allyl thiocyanates into allyl isothiocyanates (Yen and Wei, 1993).

Among the glucosinolate breakdown products, allyl isothiocyanate is mainly responsible for the pungent taste and odor of mustard seeds or processed products. This compound is formed by the action of the myrosinase enzyme on the glucosinolate compound, sinigrin (2-allyl-glucosinolates) (Eib et al., 2020). Isothiocyanates such as allyl isothiocyanate and indole-3-carbinol have been reported to exhibit antitumor activity as evidenced by multiple cancer cell-line studies (including human lung cancer). These bioactive components mainly modulate the cell cycle (G2/M phase), exhibit free-radical scavenging activity, and induce mitochondria-mediated programmed cell death (Wu et al., 2009). Sinapic acid, SA (3,5-dimethoxy-4-hydroxycinnamic acid) is a small hydroxycinnamic acid derivative that exhibits antitumor activities including the inhibition of inflammatory cytokines (Eroglu et al., 2018). The cytotoxic activity of SA suppresses the proliferation of human breast and colon cancer cells (Hudson et al., 2000). Sinapic acid also induces apoptosis in human cervical adenocarcinoma cells (Senawong et al., 2013). In fact,  $\gamma$ -linolenic acid too has been reported to inhibit colon cancer cells (MDA-MB-435) (Rose et al., 1995) as well as breast cancer cells (MCF7) (Jiang et al., 1998). Indole-3-carbinol inhibits multiple oncogenic signaling pathways and arrests the aggressiveness of neoplastic cells (Ahmed et al., 2020).

Indole-3-carbinol modulates gut microbiota by exhibiting antimicrobial activity against pathogenic bacteria as well as promoting anti-inflammatory butyrate production that helps control colitis (Busbee et al., 2020). Studies have also reported the antibacterial activity of indole-3-carbinol against *Citrobacter rodentium*, a major pathogen responsible for acute intestinal inflammation (Wu et al., 2020). Allyl isothiocyanate in both liquid and vapor forms exhibits potent bactericidal and bacteriostatic activities against food-borne pathogenic agents such as *Salmonella* Montevideo, *Escherichia coli* O157:H7, and *Listeria monocytogenes* Scott A (Lin et al., 2000). The natural bioactive compound,  $\beta$ -sitosterol possesses several pharmacological properties and is known for its antimicrobial, antitumorogenic, and immunostimulatory activities (Babu and Jayaraman, 2020). In particular,  $\beta$ -sitosterol attenuates diet-induced inflammatory reactions and can modulate gut microbiota in high-grain diet-induced dysbiosis (Xia et al., 2020). Similarly, another bioactive

compound sinapine also stimulates the prevalence of probiotics, such as Lactobacillaceae, Akkermansiaceae, and Blautia as well as inhibits the expression of inflammatory factors (Li et al., 2019). Our results reinforce the utility of traditional mustard seed fermented products as potential prebiotics for modulating the gut microbiota and ameliorating multiple clinical conditions.

## Conclusion

In Assam, *Kahudi* and *Kharoli* are consumed in small quantities as side dishes to the main meal and are cherished due to their unique organoleptic characteristics, nutritional composition, and perceived health benefits. These local and artisanal condiments are slowly and steadily finding their way from household kitchens to urban platters. Of late, new cottage food industries are growing around these food items catering to the needs of food enthusiasts. A better understanding of technical know-how, fermentation process, and nutritional and health benefits will surely help in the resurgence of mustard seed fermented products in and around Assam. Fermentation of *Kahudi* and *Kharoli* is spontaneous and uncontrolled, which increases the likelihood of product variations. Therefore, this preliminary study was undertaken to gain insights into the array of microorganisms and metabolites associated with the traditional fermentation of mustard seeds. Our integrated LC-MS-based metabolomics and NGS-based microbial diversity study revealed the dominance of endospore-forming and proteolytic *Bacillus* in *Kahudi* and *Lysinibacillus* in *Kharoli* samples as well as the abundance of many therapeutically-important bioactive metabolites such as sinapine, indole-3-carbinol,  $\gamma$ -linolenic acid,  $\beta$ -sitosterol acetate, 3-butylene glucosinolate, erucic acid, etc. in the two food products. We hope that further studies on the isolation of dominant microbiota toward the development of probiotic starter culture and enrichment of selective bioactive components will pave way for the wide commercialization of these ethnic food products.

## Data availability statement

The datasets presented in this study can either be found in online repositories or will be available on conditional request. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

## Author contributions

MB: conceptualization, funding acquisition, project administration, and supervision. SB, SD, and KD: data curation. DH, SD, and NS: formal analysis. MB and SB: investigation and writing—review and editing. MB, SB, DS, and KD: methodology. MB, US, and DS: resources. DH and NS: software. SB, DH, and

GG: validation and visualization. SB, MB, and NS: roles and writing—original draft. All authors contributed to the article and approved the submitted version.

## Acknowledgments

All authors are grateful to Prof. Bidyut Kumar Sarmah, Director, DBT—North East Centre for Agricultural Biotechnology, AAU, Jorhat, and Prof. Mahendra Kumar Modi, Head, Department of Agriculture Biotechnology for their timely suggestions and guidance during this study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships

## References

- Ahmed, A. G., Hussein, U. K., Ahmed, A. E., Kim, K. M., Mahmoud, H. M., Hammouda, O., et al. (2020). Mustard seed (*Brassica nigra*) extract exhibits antiproliferative effect against human lung cancer cells through differential regulation of apoptosis, cell cycle, migration, and invasion. *Molecules* 25, 2069. doi: 10.3390/molecules25092069
- Al-Jasas, F. M., and Al-Jasser, M. S. (2012). Chemical composition and fatty acid content of some spices and herbs under Saudi Arabia conditions. *Sci. World J.* 2012, 1–5. doi: 10.1100/2012/859892
- Babu, S., and Jayaraman, S. (2020). An update on  $\beta$ -sitosterol: a potential herbal nutraceutical for diabetic management. *Biomed. Pharmacother.* 131, 110702. doi: 10.1016/j.biopha.2020.110702
- Busbee, P. B., Menzel, L., Alrafas, H., Dopkins, N., Becker, W., Miranda, K., et al. (2020). Indole-3-carbinol prevents colitis and associated microbial dysbiosis in an IL-22-dependent manner. *JCI Insight*, 5. doi: 10.1172/jci.insight.127551
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13, 581–583. doi: 10.1038/nmeth.3869
- Cartea, M. E., Francisco, M., Soengas, P., and Velasco, P. (2010). Phenolic compounds in *Brassica* vegetables. *Molecules* 16, 251–280. doi: 10.3390/molecules16010251
- Chukeatirote, E. (2015). Thua nao: Thai fermented soybean. *J. Ethnic Foods* 2, 115–118. doi: 10.1016/j.jef.2015.08.004
- Clarke, D. B. (2010). Glucosinolates, structures and analysis in food. *Anal. Methods* 2, 310. doi: 10.1039/b9ay00280d
- Cosgrove, D. (1963). The isolation of myoinositol pentaphosphates from hydrolysates of phytic acid. *Biochem. J.* 89, 172–175. doi: 10.1042/bj0890172
- Dakwa, S., Sakyi-Dawson, E., Diako, C., Annan, N. T., and Amoa-Awua, W. K. (2005). Effect of boiling and roasting on the fermentation of soybeans into dawadawa (soy-dawadawa). *Int. J. Food Microbiol.* 104, 69–82. doi: 10.1016/j.ijfoodmicro.2005.02.006
- Eib, S., Schneider, D. J., Hensel, O., and Seuß-Baum, I. (2020). Relationship between mustard pungency and allyl-isothiocyanate content: A comparison of sensory and chemical evaluations. *J. Food Sci.* 85, 2728–2736. doi: 10.1111/1750-3841.15383
- Eroglu, C., Avcı, E., Vural, H., and Kurar, E. (2018). Anticancer mechanism of Sinapic acid in PC-3 and LNCaP human prostate cancer cell lines. *Gene* 671, 127–134. doi: 10.1016/j.gene.2018.05.049
- Fenwick, G. R., Heaney, R. K., Mullin, W. J., and VanEtten, C. H. (1983). Glucosinolates and their breakdown products in food and food plants. *C R C Crit. Rev. Food Sci. Nutr.* 18, 123–201. doi: 10.1080/10408398209527361
- Ghosh, S., and Das, A. P. (2018). Metagenomic insights into the microbial diversity in manganese-contaminated mine tailings and their role in biogeochemical cycling of manganese. *Sci. Rep.* 8, 8257. doi: 10.1038/s41598-018-26311-w
- Goswami, G., Bora, S. S., Parveen, A., Chandra Boro, R., and Barooah, M. (2017). Identification and functional properties of dominant lactic acid bacteria isolated from Kahudi, a traditional rapeseed fermented food product of Assam, India. *J. Ethnic Foods* 4:1425–1432. doi: 10.1016/j.jef.2017.08.008
- Hanschen, F. S., Lamy, E., Schreiner, M., and Rohn, S. (2014). Reactivity and stability of glucosinolates and their breakdown products in foods. *Angewandte Chem. Int. Edn.* 53, 11430–11450. doi: 10.1002/anie.201402639
- Holst, B., and Williamson, G. (2004). A critical review of the bioavailability of glucosinolates and related compounds. *Nat. Prod. Rep.* 21, 425. doi: 10.1039/b204039p
- Horwitz, W., Latimer, G. W., and International A (2005). *Official Methods of Analysis of AOAC International, 8th edn*, ed M. Gaithersburg. Rockville, MD: AOAC International.
- Hudson, E. A., Dinh, P. A., Kokubun, T., Simmonds, M. S., and Gescher, A. (2000). Characterization of potentially chemopreventive phenols in extracts of brown rice that inhibit the growth of human breast and colon cancer cells. *Cancer Epidemiol. Biomark. Prevent. A Publ. Am. Assoc. Cancer Res. Cospons. Am. Soc. Prevent. Oncol.* 9, 1163–1170.
- Jaime, R., Alcántara, J. M., Manzaneda, A. J., and Rey, P. J. (2018). Climate change decreases suitable areas for rapeseed cultivation in Europe but provides new opportunities for white mustard as an alternative oilseed for biofuel production. *PLoS ONE* 13, e0207124. doi: 10.1371/journal.pone.0207124
- Jiang, W. G., Bryce, R. P., Horrobin, D. F., and Mansel, R. E. (1998). Gamma-Linolenic acid blocks cell cycle progression by regulating phosphorylation of p27kip1 and p57kip2 and their interactions with other cycle regulators in cancer cells. *Int. J. Oncol.* 13, 611–618. doi: 10.3892/ijo.13.3.611
- Kalwasińska, A., Jankiewicz, U., Felföldi, T., Burkowska-But, A., and Swiontek Brzezinska, M. (2018). Alkaline and halophilic protease production by *Bacillus luteus* H11 and its potential industrial applications. *Food Technol. Biotechnol.* 56:553. doi: 10.17113/ftb.56.04.18.5553
- Katoch, R. (2011). “Methods for nutritional quality evaluation of food materials.” In: *Analytical Techniques in Biochemistry and Molecular Biology*. New York, NY: Springer, pp. 251–322. doi: 10.1007/978-1-4419-9785-2\_13

that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fsufs.2022.1006573/full#supplementary-material>

- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucl. Acids Res.* 41, e1–e1. doi: 10.1093/nar/gks808
- Kumar, S., Chauhan, J. S., and Kumar, A. (2010). Screening for erucic acid and glucosinolate content in rapeseed-mustard seeds using near infrared reflectance spectroscopy. *J. Food Sci. Technol.* 47, 690–692. doi: 10.1007/s13197-010-0120-3
- Li, Y., Li, J., Su, Q., and Liu, Y. (2019). Sinapine reduces non-alcoholic fatty liver disease in mice by modulating the composition of the gut microbiota. *Food Funct.* 10, 3637–3649. doi: 10.1039/C9FO00195F
- Lin, C.-M., Preston, J. F., and Wei, C.-I. (2000). Antibacterial mechanism of Allyl isothiocyanate†. *J. Food Protect.* 63, 727–734. doi: 10.4315/0362-028X-63.6.727
- Liu, X., Fang, F., Xia, X., Du, G., and Chen, J. (2016). Stability enhancement of urethanase from *Lysinibacillus fusiformis* by site-directed mutagenesis. *Sheng wu gong cheng xue bao = Chine. J. Biotechnol.* 32, 1233–1242. doi: 10.13345/jf.cjb.150527
- Manchanda, S. C., and Passi, S. J. (2016). Selecting healthy edible oil in the Indian context. *Indian Heart J.* 68, 447–449. doi: 10.1016/j.ihj.2016.05.004
- Mazumder, A., Dwivedi, A., and du Plessis, J. (2016). Sinigrin and its therapeutic benefits. *Molecules* 21, 416. doi: 10.3390/molecules21040416
- Narbad, A., and Rossiter, J. T. (2018). Gut glucosinolate metabolism and isothiocyanate production. *Mol. Nutr. Food Res.* 62, 1700991. doi: 10.1002/mnfr.201700991
- Nicolescu, T. O. (2017). “Interpretation of mass spectra,” in *Mass Spectrometry*. Rang-Du-Fliers: Intech. doi: 10.5772/intechopen.68595
- Nurudeen, A. O., and Chimezie Okorie, P. (2019). “African fermented food condiments: microbiology impacts on their nutritional values,” In *Frontiers and New Trends in the Science of Fermented Food and Beverages*. London: IntechOpen. doi: 10.5772/intechopen.83466
- Oguntoyinbo, F. A., Fusco, V., Cho, G.-S., Kabisch, J., Neve, H., Bockelmann, W., et al. (2016). Produce from Africa’s gardens: potential for leafy vegetable and fruit fermentations. *Front. Microbiol.* 0, 981. doi: 10.3389/FMICB.2016.00981
- Orsavova, J., Misurcova, L., Ambrozova, J., Vicha, R., and Mlcek, J. (2015). Fatty acids composition of vegetable oils and its contribution to dietary energy intake and dependence of cardiovascular mortality on dietary intake of fatty acids. *Int. J. Mol. Sci.* 16, 12871–12890. doi: 10.3390/ijms160612871
- Ouoba, L. I. I., Vouidibio Mbozo, A. B., Thorsen, L., Anyogu, A., Nielsen, D. S., Kobawila, S. C., et al. (2015). *Lysinibacillus louembei* sp. nov., a spore-forming bacterium isolated from Ntoba Mbodi, alkaline fermented leaves of cassava from the Republic of the Congo. *Int. J. Syst. Evol. Microbiol.* 65, 4256–4262. doi: 10.1099/ijsem.0.000570
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res.* 41, D590. doi: 10.1093/nar/gks1219
- Razzaq, A., Shamsi, S., Ali, A., Ali, Q., Sajjad, M., Malik, A., et al. (2019). Microbial proteases applications. *Front. Bioeng. Biotechnol.* 7:14. doi: 10.3389/fbioe.2019.00110
- Rose, D. P., Connolly, J. M., and Liu, X. (1995). Effects of linoleic acid and  $\gamma$ -linolenic acid on the growth and metastasis of a human breast cancer cell line in nude mice and on its growth and invasive capacity *in vitro*. *Nutr. Cancer* 24, 33–45. doi: 10.1080/01635589509514391
- Rouzaud, G. (2004). Hydrolysis of glucosinolates to isothiocyanates after ingestion of raw or microwaved cabbage by human volunteers. *Cancer Epidemiol. Biomark. Prevent.* 13, 125–131. doi: 10.1158/1055-9965.EPI-085-3
- Samtiya, M., Aluko, R. E., and Dhewa, T. (2020). Plant food anti-nutritional factors and their reduction strategies: an overview. *Food Prod. Process. Nutr.* 2, 6. doi: 10.1186/s43014-020-0020-5
- Sanni, A. I., Ayernor, G. S., Sakyi-Dawson, E., and Sefa-Dedeh, S. (2000). Aerobic spore-forming bacteria and chemical composition of some Nigerian fermented soup condiments. *Plant Foods Hum. Nutr.* 55, 111–118. doi: 10.1023/A:1008147120526
- Senawong, T., Misuna, S., Khaopha, S., Nuchadomrong, S., Sawatsitang, P., Phaosiri, C., et al. (2013). Histone deacetylase (HDAC) inhibitory and antiproliferative activities of phenolic-rich extracts derived from the rhizome of *Hydnophytum formicarum* Jack.: sinapinic acid acts as HDAC inhibitor. *BMC Compl. Alter. Med.* 13, 232. doi: 10.1186/1472-6882-13-232
- Sharafi, Y., Majidi, M. M., Goli, S. A. H., and Rashidi, F. (2015). Oil content and fatty acids composition in *Brassica* Species. *Int. J. Food Prop.* 18, 2145–2154. doi: 10.1080/10942912.2014.968284
- Sun, P., Wang, J. Q., and Zhang, H. T. (2010). Effects of *Bacillus subtilis* natto on performance and immune function of preweaning calves. *J. Dairy Sci.* 93, 5851–5855. doi: 10.3168/jds.2010-3263
- Tamang, J. P., and Nikkuni, S. (1996). Selection of starter cultures for the production of kinema, a fermented soybean food of the Himalaya. *World J. Microbiol. Biotechnol.* 12, 629–635. doi: 10.1007/BF00327727
- Trygg, J., and Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). *J. Chem.* 16, 119–128. doi: 10.1002/cem.695
- Ugwuanyi, O. J., and Okpara, N. A. (2020). “Current status of alkaline fermented foods and seasoning agents of Africa,” in *New Advances on Fermentation Processes*. London: IntechOpen. doi: 10.5772/intechopen.87052
- Vis, A. P., Rampal, G., Thind, T. S., and Arora, S. (2009). Bio-protective effects of glucosinolates—a review. *LWT Food Sci. Technol.* 42, 1561–1572. doi: 10.1016/j.lwt.2009.05.023
- Vis, P. A., and Walia, A. (2001). Beneficial effects of *Rhizopus oligosporus* fermentation on reduction of glucosinolates, fibre and phytic acid in rapeseed (*Brassica napus*) meal. *Bioresour. Technol.* 78, 309–312. doi: 10.1016/S0960-852400030-X
- Viñas, P. (1993). Rapid determination of calcium, magnesium, iron and zinc in flours using flow injection flame atomic absorption spectrometry for slurry atomization. *Food Chem.* 46, 307–311. doi: 10.1016/0308-814690125-Y
- Wang, L.-J., Li, D., Zou, L., Dong Chen, X., Cheng, Y.-Q., Yamaki, K., et al. (2007). Antioxidative activity of Douchi (a Chinese traditional salt-fermented soybean food) extracts during its processing. *Int. J. Food Prop.* 10, 385–396. doi: 10.1080/10942910601052715
- Wenzler, E., Kamboj, K., and Balada-Llasat, J.-M. (2015). Severe sepsis secondary to persistent *Lysinibacillus sphaericus*, *Lysinibacillus fusiformis* and *Paenibacillus amylolyticus* Bacteremia. *Int. J. Infect. Dis.* 35, 93–95. doi: 10.1016/j.ijid.2015.04.016
- Wu, X., Zhou, Q., and Xu, K. (2009). Are isothiocyanates potential anti-cancer drugs? *Acta Pharmacol. Sin.* 30, 501–512. doi: 10.1038/aps.2009.50
- Wu, Y., He, Q., Yu, L., Pham, Q., Cheung, L., Kim, Y. S., et al. (2020). Indole-3-carbinol inhibits *Citrobacter rodentium* infection through multiple pathways including reduction of bacterial adhesion and enhancement of cytotoxic T cell activity. *Nutrients*, 12, 917. doi: 10.3390/nu12040917
- Xia, G., Sun, J., Fan, Y., Zhao, F., Ahmed, G., Jin, Y., et al. (2020).  $\beta$ -Sitosterol Attenuates high grain diet-induced inflammatory stress and modifies rumen fermentation and microbiota in sheep. *Animals* 10, 171. doi: 10.3390/ani10010171
- Yen, G.-C., and Wei, Q.-K. (1993). Myrosinase activity and total glucosinolate content of cruciferous vegetables, and some properties of *Cabbage Myrosinase* in Taiwan. *J. Sci. Food Agric.* 61, 471–475. doi: 10.1002/jsfa.2740610415