Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

https://doi.org/10.17113/ftb.61.01.23.7796 original scientific paper

Identification of Bulgarian Sourdough Microbiota by Metagenomic Approach Using Three Commercially Available DNA Extraction Protocols

Running title: Investigation of Bulgarian Sourdough Microbiota by NGS

Ivelina Vassileva[1], Vesselin Baev[2], Galina Yahubyan[2], Elena Apostolova-Kuzova[2], Angel Angelov[3] and Miglena Koprinarova[1,4*]

[1]Institute of Molecular Biology "Acad. Roumen Tsanev", Bulgarian Academy of Sciences, Acad. G. Bonchev Str. bl. 21, 1113 Sofia, Bulgaria
[2]Department of Plant Physiology and Molecular Biology, Tzar Assen 24, University of Plovdiv, 4000 Plovdiv, Bulgaria
[3]Department of Biotechnology, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria
[4]Department of Catering and Nutrition, University of Food Technologies, 26 Maritza Blvd., 4002 Plovdiv, Bulgaria

Received: 24 May 2022
Accepted: 6 February 2023

SUMMARY

Research background. Sourdoughs are spontaneously formed, complex microbial ecosystems of various Lactic Acid Bacteria (LAB) and yeast which, by producing specific metabolites, determine the quality of the baked products. In order to design and control sourdoughs with preferred nutritional characteristics it is crucial that the LAB diversity of the product of interest be elucidated.

Experimental approach. Using the opportunities of next-generation sequencing (NGS), of the 16S rRNA of V1-V3 hypervariable gene region, we studied the microbial ecosystem of a whole grain sourdough made of *Triticum monococcum*, originated from Southwestern Bulgaria. Since the DNA extraction method is considered crucial for the accuracy of the sequencing results, as it can introduce significant differences in the analyzed microbiota, we used three different commercial kits for DNA isolation and analyzed their impact on the observed bacterial diversity.

*Corresponding author:
Phone: +359888346104
Fax: +35923723507
E-mail: mkoprinarova@abv.bg

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Results and conclusions. All three DNA extraction kits provided bacterial DNA which passed quality control and was successfully sequenced on Illumina MiSeq platform. The results received from the different DNA protocols showed variations in the microbial profiles. Alpha diversity indices (ACE, Chao, Shannon, Simpson) were also different among the three groups of results. Nevertheless, a strong dominance of phylum Firmicutes, class Bacilli, order Lactobacillales, represented mostly by family Lactobacillaceae, genus *Lactobacillus* (relative abundance of 63.11–82.28 %) and family Leuconostocaceae, genus *Weissella* (relative abundance of 3.67–36.31 %) were observed. *Lactiplantibacillus plantarum* and *Levilactobacillus brevis* with relative abundance of 16.15–31.24 % and 6.21−16.29 % respectively, were the two dominant species identified in all three DNA isolates.

Novelty and scientific contribution. The presented results give insight into the taxonomic composition of bacterial community of a specific Bulgarian sourdough. Having in mind that the sourdough is a difficult matrix for DNA isolation, on the one hand, and that there is no standardized DNA extraction protocol for this matrix, on the other hand, this pilot study aimed to give a small contribution to the future establishment and validation of such a protocol, which will allow accurate assessment of the specific microbiota of sourdough samples.

Keywords: sourdough; DNA extraction methods; V1-V316S rRNA; next-generation sequencing; metagenomics; microbiota

INTRODUCTION

In the last few years, the interest in products made of whole grain flour by sourdough fermentation in European markets has been growing fast. The use of sourdough improves the flavor, structure and stability of baked goods and ameliorates the nutritional qualities of the whole grain products by delaying flour digestibility, thus decreasing glycemic response, increasing protein digestibility and improving the bioavailability of mineral substances. It is suggested that the increased intake of sourdough products helps to improve conditions like diabetes, cardiovascular and intestinal diseases (*1*). Whole grain flour is known to contain much more vitamins, minerals, fibers, antioxidants, carotenoids, flavonoids and phenolic acids compared to refined wheat flour and sourdough fermentation increases the amount of beneficial microorganisms in whole wheat products, which improves human health (*1*). Sourdough has a dynamic ecosystem of fermentation organisms. The most common LAB species are *Fructilactobacillus sanfranciscensis* (former *Lactobacillus sanfranciscensis*), *Lactiplantibacillus plantarum* (former *Lactobacillus plantarum*), *Levilactobacillus brevis* (former *Lactobacillus brevis*), *Pediococcuspentosaceus*, *Companilactobacillus alimentarius*

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

(former *Lactobacillus alimentarius*), *Limosilactobacillus pontis* (former *Lactobacillus pontis*), *Furfurilactobacillus rossiae* (former *Lactobacillus rossiae*) (*2-4*), some species of *Leuconostoc* and *Weissella*, as well as yeasts as *Saccharomyces cerevisiae* and species of the Kazachstania clade (*5-7*). The qualitative characteristics of the fermented sourdough are defined by LAB diversity and yeast flora in it (*8-11*). The different ingredients, environmental factors and fermentation conditions define the uniqueness of each sourdough starter, characterized by a complex microbial ecosystem of interacting yeast and bacteria. Better understanding of the factors affecting sourdough microbiota provides opportunity for selection of starter cultures and fermentation conditions, which can improve the sourdough breadmaking process and the production of baked goods with desirable quality. Hence, our research was focused on the investigation of bacterial microbiota in traditional Bulgarian whole grain sourdough made of *Triticum monococcum* flour by means of NGS technology.

One of the most important methods of conventional microbiome analysis of food is 16S DNA sequencing, which provides insight into the microbial ecosystem of the product with a precise taxonomic resolution. The 16S rRNA gene contains nine hypervariable regions flanked by conserved sequences, which allow amplification and sequencing of target regions and taxonomic identification of the food associated bacterial species. The sequences are clustered into operational taxonomic units (OTUs) which are then compared against databases for identifications of the bacteria present in the microbiome. Recently, new methods have been developed that resolve amplicon sequence variants (ASVs) from NGS amplicon data without disadvantages that define molecular OTUs (*12-17*). ASV methods infer the biological sequences in the sample prior to the introduction of amplification and sequencing errors, and discriminate sequence variants differing by as little as one nucleotide.

NGS techniques are very efficient for studying the food microbiota, but there are some methodological features crucial for the success of the sequencing. One critical step affecting the accuracy of 16S rRNA NGS data, is the DNA extraction. The isolation of good quantity and quality DNA depends on the protocol used for a particular matrix (*18*) and affects the experimental results. That is why it is essential that an optimal extraction method should be used (*19-23*). A recommended step in DNA isolation from food matrices is the removal of background DNA of plant or animal origin that would affect the relevant sequence information. The matrix contents may also interfere with performance of molecular analysis as it may inhibit the required biochemical reactions (*22*). A potential approach to eliminate matrix components is to retrieve microbes by centrifugation (*24*) or differential centrifugation and filtration from aqueous solutions (*25*). Concerning the sourdough, the elasticity of the samples further complicates the analysis, which requires thorough homogenization of the samples.

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Different commercial kits have been used for DNA extraction for metagenome analysis from sourdough: DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) (*26*), DNeasy PowerSoil kit (Qiagen, Hilden, formerly MoBio) (*27*) as well as from other matrices: Zymo Quick-DNA Fecal/Soil Microbe Mini-prep Kit (Ozyme, Saint-Cyr-l'Ecole, France) (*28*), DNeasy Blood and Tissue kit (Qiagen, Venlo, Netherlands) (*29*). However, the choice of DNA extraction kit for isolation of bacterial DNA from sourdough is still difficult and further investigation is required to elucidate the efficiency of the commercial kits available on the market.

In this research, we chose V1-V3 hypervariable region of 16S rRNA for NGS and analysed the microbiota of Bulgarian sourdough from *T. monococcum*. For better characterization we used three different DNA extraction kits from three manufacturers namely: NucleoSpin Food kit (Macherey-Nagel, Dueren, Germany) for DNA extraction from food samples specifically, QIAamp DNA Microbiome Kit (Qiagen, Hilden, Germany) for isolation of bacterial microbiome DNA, and GENESpin kit (Eurofins Technologies, Budapest, Hungary) for isolation of high-quality DNA from food and feed samples. Two of the kits (NucleoSpin and GENESpin) are designed to extract DNA from food, while QIAamp is intended for depletion of host DNA and extraction of microbiome DNA from mixed samples. In addition, NucleoSpin and QIAamp kits were chosen because they have already been used in published metagenomic analysis of food (*30*) and other type of samples (*31*). To the best of our knowledge, the GENESpin kit has not been applied for metagenomic analysis so far.

MATERIALS AND METHODS

Sourdough

The sourdough originated from a manufacturer of typical Bulgarian bread based in Southwestern Bulgaria (Fig. S1). The whole grain flour used was from *Triticum monococcum*. The technological time for the fermentation of the sourdough was 18 hours at 22 °C and included three backsloppings. For each backslopping, 1.5 % NaCl (*m*/*V*) was added and 25 % of the sourdough was used. The sample was taken after the last backslopping and was kept and transported frozen until the DNA extraction was performed. Before extraction, the sourdough was well mechanically homogenized by adding 5 mL sterile peptone water (1 % peptone (*m*/*V*) and 0.9 % NaCl (*m*/*V*)) to 500 mg of the probe.

DNA extraction

We compared three kits to evaluate their capability to extract high quality DNA from sourdough for 16S rRNA NGS analysis. These are namely NucleoSpin Food kit (Macherey-Nagel, Dueren

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Germany) for DNA extraction from food samples specifically, QIAamp DNA Microbiome Kit (Qiagen, Hilden, Germany) for isolation of bacterial microbiome DNA form mixed samples, and GENESpin kit (Eurofins Technologies, Budapest, Hungary) for isolation of high-quality DNA from food and feed samples. These three different commercial kits were used for DNA extraction from sourdough according to the manufactures' protocols with several optimizations for the particular matrix. The names of all reagents (*e.g.* buffers, enzymes) mentioned hereafter are as provided by the respective manufacturer.

Method 1 (NucleoSpin Food Kit)

DNA isolation with Macherey-Nagel Food Kit was performed in the laboratories of Eurofins Genomics Europe Sequencing GmbH. The protocol had been previously validated. DNA was extracted from 200 mg of the starting sample material after homogenization. The NucleoSpin Food protocol consists of six steps for extraction of genomic DNA (plant, animal or bacteria) from food samples. It guarantees good recovery of small genomic DNA fragments (<1kbp) out of complex food matrices. The sample was first well homogenized. Then 550 µL of Lysis Buffer CF (NucleoSpin Food Kit) and 10 µL Proteinase K (10 mg/mL, NucleoSpin Food Kit) were added to the homogenate and incubated for 30 min at 65 °C. After centrifugation for 10 min at 10 000×*g*, to the clear supernatant (volume (*V*) 1) was added 1 *V* Buffer C4 (NucleoSpin Food Kit) and 1 *V* ethanol. For DNA binding, the sample was loaded (maximum 700 µL at a time) onto one NucleoSpin Food Column (NucleoSpin Food Kit), which was placed in a collection tube and centrifuged for 1 min at 11 000 ^{x}g. After discarding the flow-through, the remaining sample was loaded. To wash the silica membrane, 400 µL of Buffer CQW (NucleoSpin Food Kit) was added, the column was centrifuged for 1 min at 11 000×*g* and the flow-through - discarded. Afterwards, 700 µL of Buffer C5 (NucleoSpin Food Kit) was added onto the NucleoSpin Column and centrifuged for 1 min at 11 000×*g*. For the third wash 200 µL of Buffer C5 was pipetted onto the column and centrifuged for 2 min at 11 000×*g* in order to remove residual ethanol completely. To elute the DNA from the membrane, the column was placed in a new 1.5 mL microcentrifuge tube and 100 µL of preheated to 70 °C Elution Buffer CE (NucleoSpin Food Kit) was added. After 5 min incubation at room temperature, it was centrifuged for 1 min at 11 000×*g* to elute the DNA.

Method 2 (QIAamp DNA Microbiome Kit)

The QIAamp DNA Microbiome Kit provides selective isolation of bacterial DNA from mixed samples. The protocol consists of six steps where the first two steps assure differential lysis and

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

degradation of background (plant) DNA, while keeping the bacterial cells intact. The nature of the test samples required pre-suspension, which was performed in PBS (Phosphate-Buffered Saline, 1.8 mM KH2PO4, 10.0 mM Na2HPO4, 2.7 mM KCl, 137.0 mM NaCl, pH=7.5) buffer, pH=7.5 (*32*) at a ratio of 1:1 (m/V). To remove coarse particles, the suspension was centrifuged at 2000×*g* for 1 min (Eppendorf centrifuge 5418, Eppendorf AG, Hamburg, Germany). Aliquots of each sample were used for two parallel extractions. For each extraction, 500 µL Buffer AHL (QIAamp DNA Microbiome Kit) were pipetted to 1 mL of the suspension in a 2 mL tube (Eppendorf, Hamburg, Germany). The mixture was incubated for 30 min at room temperature with rotation and then centrifuged at 10 000×*g* for 10 min (Eppendorf centrifuge 5418). After removing the supernatant, 190 µL of Buffer RDD (QIAamp DNA Microbiome Kit) and 2.5 µL of Benzonase were pipetted and mixed well prior to incubation at 37 °C for 30 min at 600 rpm in a heating block (Thermomixer Comfort, Eppendorf AG, Hamburg, Germany). Following treatment with 20 µL Proteinase K (QIAamp DNA Microbiome Kit) at 56 °C for 30 min at 600 rpm, and briefly spinning the tube at slow speed, 200 µL of Buffer ATL (with Reagent DX, QIAamp DNA Microbiome Kit) was mixed well with the sample to avoid loss of sample material. With this final step, the lysis of the eukaryotic cells was completed. The third step of chemical and mechanical disruption of bacterial cells included transfer of the sample into a Pathogen Lysis Tube L for cell lysis in a Tissue Lyser LT, FastPrep24 instrument (Qiagen, Hilden, Germany) for disruption (2 times at 50Hz for 10 min). After centrifugation of the Pathogen Lysis Tube L, 40 µL Proteinase K (QIAamp DNA Microbiome Kit) was added to the supernatant and 200 µL of the Buffer APL2 (QIAamp DNA Microbiome Kit) were added by pulse vortexing. After incubation, 200 µL ethanol was added to the lysate. To remove intact non-lysed cells, the lysate was centrifuged at 2000^{x}*g* for 1 min on Eppendorf centrifuge 5418. The fourth step was the lysate adsorption to the QIAampUCP membrane. At this step, the lysates from the two parallel extractions were combined (approximate final volume 1 600 µL) and 700 µL was transferred into the QIAampUCP Mini column (QIAamp DNA Microbiome Kit) followed by centrifugation at 6000×*g* for 1 min. The rest of the lysate was loaded on the same column and centrifugated at the same conditions. The next two steps included washing with 500 µL of Buffer AW1 (QIAamp DNA Microbiome Kit), followed by centrifugation and then the addition of Buffer AW2 and centrifugation at full speed for 3 min. The QIAampUCP Mini Column (QIAamp DNA Microbiome Kit) was placed into a fresh 2 mL collection tube and centrifuged at full speed for 1 min. The last step of DNA elution was performed by adding 30 µL Buffer AVE (QIAamp DNA Microbiome Kit) onto the column, 5 min incubation at room temperature and centrifugation at 6 000×*g* for 1 min.

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Method 3 (GENESpin Kit)

The GENESpin Kit is designed for the isolation of genomic DNA from food and feed samples of plant and animal origin. The kit assures good recovery rates for small genomic DNA fragments (<1kb). This protocol consists of six steps including three washing steps such as the NucleoSpin Food protocol. The starting sample material was 200 mg, taken after previous homogenization. The second step for cell lysis was performed by adding 550 µL of GENESpin Lysis Buffer, pre-heated to 65 °C (WNB14 Memmert waterbath, Schwabach, Germany), and incubation with 10 µL Proteinase K (10 mg/mL, GENESpin Kit) for 30 min at 65 °C. 10 µL RNAse A (20 mg/mL, Eurofins Technologies, Budapest, Hungary) was added to the sample and incubation for 30 min at RT was performed as well. Following centrifugation (Eppendorf centrifuge 5418, Eppendorf AG, Hamburg, Germany), the supernatant was transferred into a new centrifuge tube (Eppendorf, Hamburg, Germany), then, one volume (*V*) of GENESpin Binding Buffer and 1 *V* ethanol (99.9%, Merck, Darmstadt, Germany) were added to the sample. The fourth step was binding the DNA onto the column matrices. A GENESpin Column was placed into a new collection tube (Eppendorf, Hamburg, Germany) and 700 µL of the sample was loaded onto the column. Afterwards, centrifugation and discarding the flow-through, the step was repeated. The fifth step was washing the sample with three washing substeps. First, 400 µL of GENESpin Wash Buffer 1 was added to the column and centrifuged, then 700 µL of GENESpin Wash Buffer 2 was added and centrifuged and the final 200 µL volume of GENESpin Wash Buffer 2 was added and centrifuged for 2 min. The last step was performed by 100 µL GENESpin Elution Buffer loaded onto the membrane, 5 min incubation and 1 min centrifugation at 11 000×*g* to elute the DNA.

For all three extracted DNA isolates, PCR and quality control were performed by Eurofins Genomics in order to check whether a PCR product could be generated. Based on this quality control, which met the requirements for amount, concentration and quality of the extracted DNA, all three isolates were subjected to next-generation sequencing.

DNA quantification

DNA quantity was determined fluorometrically with Qubit 4 Fluorometer (Invitrogen™, Thermo Fisher Scientific, Waltham, USA). DNA integrity was checked by electrophoresis on 1 % agarose gel.

Next-generation sequencing of 16S rRNA gene amplicons

The amplification of target sequences of 16S rRNA gene and NGS was performed by Eurofins Genomics, Germany. The extracted DNA was subjected to PCR amplification of 16S rRNA

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

gene sequences from the hypervariable regions V1-V3 for bacterial profiling. The amplicons were generated from the samples *via* two-step PCR protocol. PCR products were subjected to sequencing on Illumina MiSeq platform 2 × 300 bp paired-end reads.

Bioinformatics data analysis

We used cutadapt v. 1.9 program (*15*) for the process of quality filtering and trimming of the pair-end reads. Metagenomics analysis was done using the DADA2 (version 1.14) pipeline as described by Callahan and co-workers (2017), which was recommended to replace OTU with amplicon sequence variants (ASV)-based approaches (*33,34*). Quality checks were conducted; clean reads were denoised, dereplicated, and filtered for chimeras to generate ASVs. The resulting amplicon sequence variant ASV table was used for all downstream analyses, including taxonomic assignment, phylogenetic alpha diversity measurements, and differential abundance comparisons, and visualizations. Taxonomic assignment of sequence variants was performed using a combination of the functions assign Taxonomy and assign Species and was compared using the SILVA reference database (v132) (*35*). We generated within-sample microbial diversity (alpha diversity) indexes (ACE, Chao1, Shannon, Simpson) by using QIIME package (alpha_diversity.py -o alphadiv --metrics 'ace,chao1,observed_otus,shannon,simpson') (*36*). Krona charts for visualization are created from samples ASV table (*37*).

RESULTS AND DISCUSSION

DNA extraction and quantification

In this research, we investigated the microbiota of spontaneously fermented sourdough from *T. monococcum*, originated from Southwestern Bulgaria, by 16S rRNA gene NGS analysis. In order to do that, we used three commercial kits for DNA extraction; specifications are given in Table 1. The quantity and yield of the extracted DNA are presented in Table 2. The extracted DNA was with different concentrations and ranged between 29.2 and 246.8 ng/μL, the total DNA yield was between 0.88 and 24.68 μg depending on the DNA extraction method (Table 2). The differences in DNA concentration and yield could be partly explained with the methodological features of the protocols. GENESpin Kit as well as NucleoSpin do not include a step for depletion of background DNA, so the DNA obtained with these two kits also included DNA with plant origin. With the QIAamp Kit, the DNA with eukaryotic origin was significantly reduced (see below). There are recommendations in the literature for reduction of host (background) DNA in isolates for microbiome analysis (*31*). Using the same QIAamp DNA Microbiome Kit among others, Bjerre and co-authors advised reduction of host

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

DNA before 16S metagenomics analysis and proved that the protocol did not introduce taxonomic biases (*31*). Although the QIAamp Kit provided selective isolation of bacterial DNA, one drawback of using this kit was the hands-on time, which in our hands was 50 min longer in comparison with the other two kits, due to the additional step for degradation of plant cells. Moreover, the price of this kit was higher (€11.92 per sample) than the prices of the other two kits (€4.02 for NucleoSpin and €2.98 for GENESpin) (Table 1). Importantly, despite the differences, all three protocols (Method 1, Method 2, Method 3, see Materials and Methods) provided DNA with quality and quantity which met the requirements of the NGS quality control. Our results confirmed previously published data which showed that the success of the 16S rRNA gene sequencing analysis is generally independent of the concentrations of DNA in the samples (*31, 38)*. Accordingly, with the QIAamp kit we obtained a lower DNA yield (Table 2), which was expected considering that this kit, unlike the other two, eliminates plant DNA, which is much more strongly represented in the sample compared to bacterial DNA. However, with the same QIAamp isolate, we achieved the highest number of reads (Table 3) and significantly increased sequencing depth (Fig. 1).

Table 1 Table 2

Metagenomics data analysis

We chose the V1-V3 region from 16S rRNA gene for sequencing. In total, 210 555 reads were analysed from the three DNA isolates of the sourdough. The number of reads in NG-26064 isolate (Method 1) was 52474, in NG-260665 isolate (Method 2) was 126 553, and in NG-26434 isolate (Method 3) – 31 528 (Table 3). The observed richness, presented by identified amplicon sequence variants (ASVs), was reported as follows for the particular DNA isolates: 45 ASVs for NG-26064, 26 ASVs for NG-26065, and 27 ASVs for NG-26434. The total number of ASVs found was 98 (Table 3).In agreement with other reports (*38*), our results showed that the DNA isolation method might influence the number of observed ASVs.

The effect of the extraction methods on alpha-diversity was assessed by Chao1, ACE, Shannon and Simpson indices. For evaluation of the species richness, ACE and Chao1 indices were estimated (*39*). Lower values for both indices were calculated for NG-26065 (QIAamp) and NG-26434 (GENESpin) isolates. Higher richness indices were calculated for NG-26064 (NucleoSpin) isolate (Table 3). The diversity indices Shannon and Simpson (*39*) were also calculated for each isolate. The lowest diversity was obtained with the QIAamp Kit, intermediate diversity was obtained with the GENESpin Kit, whereas the highest diversity was achieved again by the NucleoSpin Kit. Therefore, the NucleoSpin Kit differs from the other two kits with higher Chao1/ACE richness and

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Shannon/Simpson diversity. This correlated with the higher number of ASVs in this isolate (Table 3). In agreement with data from the literature, our results showed that the different DNA extraction methods (Table 1) might influence the bacterial species richness and diversity (*40*).

We generated rarefaction plot to investigate whether the sequencing depth was sufficient to assess the microbiota in all three isolates. The calculated sequencing depth was based on alpha diversity analyses (Fig. 1). The three observed rarefaction curves each reached a plateau, suggesting that the ASVs presented in all isolates were almost completely determined and a potential increase of number of reads would not significantly contribute to the number of ASVs. QIAamp isolate had the highest sequencing depth, whereas GENESpin isolate – the lowest. The comparative analysis highlighted NucleoSpin as the best among the three kits for extraction of bacterial DNA for NGS from sourdough, since at sufficient sequencing depth the number of ASVs was much bigger compared to those achieved with the other two kits. Nevertheless, the rarefaction curve analysis showed that the depth of 16S rRNA gene sequencing was adequate in all three isolates, which allowed observation of the bacterial community of the sourdough.

Table 3 Figure 1

Taxonomic profile of the sourdough sample.

In agreement with data from the literature (*41*) the 16S rRNA sequence analysis revealed that Gram–positive bacteria were dominant in the investigated sourdough compared with Gram-negative. The dominant taxon at Phylum level was Firmicutes (87.67 %); established in 86 % in NG-26064 isolate, 100 % in NG-26065, and 77 % in NG-26434 isolate. Bacilli was the most dominant encountered class, 86 % in NG-26064 isolate, 100 % in NG-26065 isolate, and 77 % in NG-26434 isolate. Lactobacillales was found as the most dominant order affiliated with class Bacilli, present in 86 % in NG-26064 isolate, 100 % in NG-26065 isolate, and 77 % in NG-26434 isolate. At the family level, Lactobacillaceae was found as dominant, represented as follows: NG-26064 – 82 %, NG-26065 – 63 %, and NG-26434 – 70 %. The sub-dominant family found was Leuconostocaceae, represented in 4 % in NG-26064, 36 % in NG-26065, and 7 % in NG-26434 GENESpin (see also Supplementary data).

The NGS analysis revealed the following ten genera in the analyzed sourdough: *Lactobacillus, Weissella, Modestobacter, Microbacterium, Leuconostoc, Corynebacterium, Pontibacter, Paracoccus, Curtobacterium,* and *Acinetobacter* (Fig. 2). Among them *Lactobacillus* was the dominant genus, represented in 82.28 % in NG-26064 isolate, 63.11 % in NG-26065 isolate, and 70.08 % in NG-26434 isolate (Fig. 2). It is important to clarify that in the present work, due to the data

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

obtained from the software (DADA2), we used the old classification of genus *Lactobacillus,* before its reclassification into 25 genera and the union of Lactobacillaceae and Leuconostocaceae published by Zheng and co-authors in 2020 (*42*). *Weissella* was the sub-dominant genus in all three isolates, represented in 3.67 % in DNA extracted with NucleoSpin, 36.31 % of QIAamp isolate, and in 6.69 % of GENESpin isolate (Fig. 2). The other genera were detected at very low abundance, less than 0.5 %. Having in mind that DADA2 software provides very restrictive data from the 16S rRNA gene NGS analysis, we considered that although the percentages are low, the presence of these genera in the DNA isolates is important to note. Accordingly, the genus *Leuconostoc* was identified in the three isolates, in NG-26064 in 0.24 %, NG-26065 in 0.09 %, and in NG-26434 in 0.18 %. Genus *Acinetobacter* was represented in 0.04 % in NG-26064 and in 0.06 % in NG-26434. We did not detect this genus by means of QIAamp Kit (NG-26065). The other abovementioned genera were found only in NG-26064 isolate as follows: *Modestobacter* – 0.30 %, *Microbacterium* – 0.29 %*, Corynebacterium* – 0.13 %*, Pontibacte r*- 0.13 %, *Paracoccus* - 0.12 %, and *Curtobacterium* - 0.06 %.

In agreement with data from the literature (*43*), facultative heterofermentative *Lactiplantibacillus plantarum* was found in codominance with obligately heterofermentative LAB in the investigated sourdough. *L. plantarum* was present in 16.15 % of NG-26064 (NucleoSpin), in 16.28 % of NG-26434 (GENESpin), and in 31.24 % of NG-26065 (QIAamp). Obligately heterofermentative *Levilactobacillus brevis* was found with relative abundance of 14.46 % and 16.29 % of NG-26064 and NG-26434, respectively, and 6.21 % of NG-26065. *Limosilactobacillus fermentum,* also obligately heterofermentative LAB, was detected in NG-26064 and NG-26434 isolates in concentrations of 0.06 % and 0.19 %, respectively. In addition, *Acinetobacter lwoffii* was identified only in NG-26064, with relative abundance of 0.04 %.

Some substantial differences were observed between results obtained with the QIAamp Kit and both NucleoSpin Kit and GENESpin Kit. Genus *Weissella* was identified with a relative abundant of 36.31 % with QIAamp, which was significantly higher than the percent identified by both NucleoSpin Kit (3.67 %) and GENESpin Kit (6.69 %) kits (Fig. 2). On the contrary, genus *Lactobacillus* was identified in the lowest percent (63.11 %) in NG-26065 (QIAamp Kit), compared with 82.28 % and 70.08 % in NG-26064 (NucleoSpin Kit) and NG-26434 (GENESpin Kit), respectively. *L. plantarum* was present in 31.24 % in NG-26065 (QIAamp) isolate compared to 16 % (16.15 % and 16.28 %) of NucleoSpin and GENESpin isolates. However, *L. brevis* was identified in the lowest percent in QIAamp isolate (6.21 %) compared to 14.46 % and 16.29 % in NG-26064 (NucleoSpin Kit) and NG-260434 (GENESpin Kit) isolates, respectively. Therefore, similarity in the taxonomic composition of bacterial community in NG-26064 (NucleoSpin Kit) and NG-26434 (GENESpin Kit) isolates was

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

identified in contrast to NG-26065 (QIAamp Kit) isolate, which differs mainly in the proportional distribution of the established ASVs. The results obtained with the QIAamp Kit can be only partly explained by the significant removal of background DNA in the particular isolate. The relative abundance of background DNA in NG-26065 (QIAamp Kit) isolate was 0.48 %, in NG-26064 (NucleoSpin Kit) isolate - 12.74 %, and in NG-26434 (GENESpin Kit) - 22.99 %. Having in mind that the presence of a large amount of background DNA can negatively affect the 16S rRNA gene sequence analysis, it seems that the GENESpin Kit is less suitable for microbiome profiling of sourdough. On the other hand, the 16S rRNA gene libraries prepared from metagenomic DNA extracted with the GENESpin Kit had better diversity and evenness (Shannon and Simpson) indices (Table 3) compared to NG-26065 (QIAamp) isolate.

In order to illustrate the information from the NGS analysis, a Venn diagram was used to depict how the results obtained from the three kits (the shared and unique ASVs) relate to each other against an overall ASVs dataset (Fig. 3). The three isolates shared only 19 of the total 98 recovered ASVs, indicating that the DNA extraction methods significantly influenced the analysis of microbiota community structure. Hence, the overlap in 19.34 % of ASVs showed that the different extraction methods captured different bacteria in the tested sourdough. The shared ASVs between NucleoSpin and QIAamp methods were 22.45 %, between NucleoSpin and GENESpin - 21.43 %, and between GENESpin and QIAamp - 19.34 %. Other authors also reported that different DNA extraction methods affected the recovery of ASVs, hence evaluation of microbiota of naturally fermented foods (*40*).

Figure 2 Figure 3

The data obtained from NGS analysis of the three isolates together provided insight into the microbiota composition of the investigated Bulgarian sourdough and the contribution of the established bacteria for the characteristics of the final product. It is well known that the lactic acid bacteria metabolic activities improves the bread structure and elongates shelf-life due to acidification and antimicrobial compound production (*44*). According to data from the literature, in sourdoughs, *Lactobacillus* strains are more frequently found than *Leuconostoc* and *Weissella* species. Our results confirmed *Lactobacillus* dominance (63.11 %-82.28 %) and sub-dominant presence of *Weissella* (3.67 %-36.31 %), followed by *Leuconostoc* with significantly less relative abundance (0.09 %-0.24 %) in all three isolates of the investigated Bulgarian sourdough. Other authors also showed the important role of representatives of the two genera *Lactobacillus* and *Weissella* in the fermentation processes of sourdoughs (*45*). LAB found in sourdough mainly belong to genus *Lactobacillus* (*43*). In confirmation of our results, there are published data that *L. plantarum* and *L. brevis* are dominant LAB in sourdoughs as well as that *L. plantarum* is generally codominant with obligately heterofermentative

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

LAB, including *L. brevis* (*46*). Homofermentative and heterofermentative LAB perform different functions and contribute to the final product quality (*1*). Heterofermentative LAB mainly produce ethyl acetate with some alcohols and aldehydes, and homofermentative LAB produce diacetyl and other carbonyls. Among different metabolites produced by LAB, lactic and acetic acids are thought to be the main organic antifungal compounds (*46*). *L. plantarum*, produces a high amount of acetic acid in sourdough as well as bacteriocins, which have an inhibiting effect against pathogens like *Bacillus subtilis* as well as antifungal activity (*43*). Preservative effect, which prolongs the shelf-life of the products, was also reported for the well-adapted LAB *L. fermentum* due to its anti-mould activity (*47*). Furthermore, dough acidification has been shown to have significant effects on the quality characteristics of bread such as texture and volume. The lactic acid mainly produced by heterofermentative LAB may be responsible for more elastic gluten structure. The lactic and acetic organic acids are also responsible for creating different taste and odor in sourdough (*46*). *L. brevis* showed higher organic acid production compared to *L. plantarum*. In addition, *L. plantarum* are known to produce a wide range of volatile compounds (*44*). *L. plantarum* and *L. brevis* have beneficial effects in bread organoleptic properties (volume, crumb texture, unique flavour) (*46*). Therefore, the synergetic effect of the LAB strains in the sourdough had a significant role in improvement of volume, texture, staling rate, and microbial shelf-life. The heterofermentative cocci *Weissella* and *Leuconostoc,* also found in the investigated sourdough, can be important for growth association with lactobacilli (*1*). Recently, in a study comparing starters composed of different combinations of dominant bacteria, the best result was obtained with a culture containing *Lactiplantibacillus plantarum* 2MI8 and exopolysaccharide (EPS)-producing *Weissella confusa/cibaria*6PI3 strains (*48*).

The presence of *Acinetobacter lwoffii* in the investigated sourdough, found in a very low count only in NG-26064 isolate, could be explained by contamination (from the environment). *Acinetobacter lwoffii* like other species from genus Acinetobacter are common in marine fish and water microbiome and are considered as pathogens. *Acinetobacter lwoffii* was found in some traditionally fermented foods like traditional sour cream from Russia (*49*) as well as in vegetables like Chinese chive (*50*).

CONCLUSIONS

In this work we investigated the taxonomic composition of the bacterial community of a traditional sourdough originated from Southwestern Bulgaria using three commercial kits for DNA isolation. The16S rRNA NGS data confirmed that the choice of DNA extraction protocol is a key factor in metagenomic analysis due to the variations generated in the recovery of ASVs by different methods of DNA isolation. The data also revealed that each of the tested kits could be used for DNA isolation

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

and estimation of the bacterial community of sourdough, although, the best results were achieved in this study with the NucleoSpin kit. The diversity in bacterial community profiles recovered by different methods emphasized the necessity of selection and validation of a standard protocol for isolation of bacterial DNA for NGS analysis of fermented foods, in particular sourdough.

ACKNOWLEDGEMENTS

We greatly acknowledge the provided service and all given information by Eurofins Technologies, Germany.

FUNDING

This research was funded by the Bulgarian National Science Fund, grant number KP-06-N-36/3 by the Bulgarian Ministry of Education and Science under the National Research Programme "Healthy Foods for a Strong Bio-Economy and Quality of Life" approved by DCM # 577/17.08.2018.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

SUPPLEMENTARY MATERIALS

All supplementary materials are available at: http://web.uni-plovdiv.bg/vebaev/16s_metagenomics/Suppl.material.zip

AUTHORS' CONTRIBUTION

Ivelina Vassileva wrote and edited the manuscript, conducted experiments, analyzed data and collected references, Vesselin Baev did the bioinformatics analysis, interpreted and visualized the data and edited the manuscript, Galina Yahubyan designed experiments and their methodology, did investigation and edited the manuscript, Elena Apostolova-Kuzova designed experiments and their methodology, did investigation and edited the manuscript, Angel Angelov conceptualized the study and its methodology and obtained funding to conduct the research, Miglena Koprinarova conceptualized and designed the study, analyzed, interpreted and visualizedthe data, wrote and edited the manuscript.

All authors have read and agreed to the published version of the manuscript.

ORCID ID

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Ivelina Vassileva -https://orcid.org/0000-0001-6457-1329

VesselinBaev- https://orcid.org/0000-0002-5224-9145

Galina Yahubyan: https://orcid.org/0000-0003-4278-5275

Elena Apostolova-Kuzova -https://orcid.org/0000-0003-1094-8009

Angel Angelov- https://orcid.org/0000-0002-3239-4724

Miglena Koprinarova - https://orcid.org/0000-0002-1926-7974

REFERENCES

1. Lau SW, Chong AQ, Chin NL, Talib RA, Basha RK. Sourdough microbiome comparison and benefits. Microorganisms. 2021;9(7):1355. http//doi.org/10.3390/microorganisms9071355

2. Corsetti A, Settanni L, van Sinderen D, Felis GE, Dellaglio F, Gobbetti M. *Lactobacillus rossii sp. nov*., isolated from wheat sourdough. Int J Syst Evol Microbiol. 2005;55(Pt 1):35-40. http://doi.org/10.1099/ijs.0.63075-0

3. Ganzle MG, Loponen J, Gobbetti M. Proteolysis in sourdough fermentations: mechanisms and potential for improved bread quality. Trends Food Sci Technol. 2008;19(10):513-21. http://doi.org/10.1016/j.tifs.2008.04.002

4. Vogel RF, Knorr R, Muller MR, Steudel U, Ganzle MG, Ehrmann MA. Non-dairy lactic fermentations: the cereal world. Antonie Van Leeuwenhoek. 1999;76(1-4):403-11. https://doi.org/10.1023/A:1002089515177

5. Ganzle M, Ripari V. Composition and function of sourdough microbiota: from ecological theory to bread quality. Int. J. Food Microbiol. 2016;239:19-25. http://doi.org/10.1016/j.ijfoodmicro.2016.05.004

6. Siepmann FB, Ripari V, Waszczynskyj N, Spier MR. Overview of sourdough technology: from production to marketing. Food Bioproc Tech. 2018;11(2):242-70. http://doi.org/10.1007/s11947-017-1968-2

7. Weckx S, Van Kerrebroeck S, De Vuyst L. Omics approaches to understand sourdough fermentation processes. Int. J. Food Microbiol. 2019;302:90-102. http://doi.org/10.1016/j.ijfoodmicro.2018.05.029

8. Ripari V, Bai YP, Ganzle MG. Metabolism of phenolic acids in whole wheat and rye malt sourdoughs. Food Microbiol. 2019;77:43-51. http://doi.org/10.1016/j.fm.2018.08.009

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

9. Milanovic V, Osimani A, Garofalo C, et al. Selection of cereal-sourced lactic acid bacteria as candidate starters for the baking industry. PLoS One. 2020;15(7): e0236190. http://doi.org/10.1371/journal.pone.0236190

10. Cizeikiene D, Jagelaviciute J, Stankevicius M, Maruska A. Thermophilic lactic acid bacteria affect the characteristics of sourdough and whole-grain wheat bread. Food Biosci. 2020;38(2):100791. http://dx.doi.org/10.1016/j.fbio.2020.100791

11. Ma S, Wang Z, Guo X, et al. Sourdough improves the quality of whole-wheat flour products: mechanisms and challenges-a review. Food Chem. 2021;360:130038. http://doi.org/10.1016/j.foodchem.2021.130038

12. Eren AM, Maignien L, Sul WJ, et al. Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol. 2013;4(12):1111-1119. http://doi.org/10.1111/2041-210X.12114

13. Tikhonov M, Leach RW, Wingreen NS. Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. ISME J. 2015;9(1):68-80. http://doi.org/10.1038/ismej.2014.117

14. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. ISME J. 2015;9(4):968-79. http://doi.org/10.1038/ismej.2014.195

15. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3. http://doi.org/10.1038/nmeth.3869

16. Amir A, McDonald D, Navas-Molina JA, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems. 2017;2(2):e00191-16. http://doi.org/10.1128/mSystems.00191-16

17. Edgar RC. UNOISE2: improved error-correction for Illumina 16 S and ITS amplicon sequencing. bioRxiv 081257. 2016. http://doi.org/10.1101/081257

18. Cao Y, Fanning S, Proos S, Jordan K, Srikumar S. A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies. Front Microbiol. 2017;8:1829. http://doi.org/10.3389/fmicb.2017.01829

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

19. De Filippis F, Parente E, Ercolini D. Metagenomics insights into food fermentations. Microb Biotechnol. 2017;10(1):91-102.
http://doi.org/10.1111/1751-7915.12421

20. Bag S, Saha B, Mehta O, et al. An improved method for high quality metagenomics DNA extraction from human and environmental samples. Sci. Rep. 2016;6:26775.
https://doi.org/10.1038/srep26775

21. Cottier F, Srinivasan KG, Yurieva M, et al. Advantages of meta-total RNA sequencing (MeTRS) over shotgun metagenomics and amplicon-based sequencing in the profiling of complex microbial communities. Npj Biofilms and Microbiomes. 2018;4:2.
https://doi.org/10.1038/s41522-017-0046-x

22. de Boer P, Caspers M, Sanders JW, et al. Amplicon sequencing for the quantification of spoilage microbiota in complex foods including bacterial spores. Microbiome. 2015;3:30.
https://doi.org/10.1186/s40168-015-0096-3

23. Panek M, Paljetak HC, Baresic A, et al. Methodology challenges in studying human gut microbiota - effects of collection, storage, DNA extraction and next generation sequencing technologies. Sci. Rep. 2018;8(1):5143.
http://doi.org/10.1038/s41598-018-23296-4

24. Koprinarova M. Methods to improve molecular detection of *Salmonella* in complex herbal matrices containing inhibitors. J Food Prot. 2021;84(8):1309-14.
http://doi.org/10.4315/JFP-20-262

25. Van Belkum A, Broadwell D, Lovern D, Petersen L, Weinstock G, Dunne WM, Jr. Proteomics and metabolomics for analysis of the dynamics of microbiota. Expert Rev Proteomics. 2018;15(2):101-4.
http://doi.org/10.1080/14789450.2018.1421073

26. Comasio A, Verce M, Van Kerrebroeck S, De Vuyst L. Diverse microbial composition of sourdoughs from different origins. Front Microbiol. 2020;11:1212.
http://doi.org/10.3389/fmicb.2020.01212

27. Reese AT, Madden AA, Joossens M, Lacaze G, Dunn RR. Influences of ingredients and bakers on the bacteria and fungi in sourdough starters and bread. mSphere. 2020;5(1): e00950-19.
http://doi.org/10.1128/mSphere.00950-19

28. Bokulich NA, Bergsveinson J, Ziola B, Mills DA. Mapping microbial ecosystems and spoilage-gene flow in breweries highlights patterns of contamination and resistance. Elife. 2015;4:e04634.
https://doi.org/10.7554%2FeLife.04634

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

29. Menezes LAA, Sardaro MLS, Duarte RTD, et al. Sourdough bacterial dynamics revealed by metagenomic analysis in Brazil. Food Microbiol. 2020;85:103302.
http://doi.org/10.1016/j.fm.2019.103302

30. Butler C, Matsumoto A, Rutherford C, Lima HK. Comparison of the effectiveness of four commercial DNA extraction kits on fresh and frozen human milk samples. Methods Protoc. 2022;5(4):63.
http://doi.org/10.3390/mps5040063

31. Bjerre RD, Hugerth LW, Boulund F, Seifert M, Johansen JD, Engstrand L. Effects of sampling strategy and DNA extraction on human skin microbiome investigations. Sci Rep. 2019;9(1):17287.
http://doi.org/10.1038/s41598-019-53599-z

32. Lhomme E, Lattanzi A, Dousset X, et al. Lactic acid bacterium and yeast microbiotas of sixteen French traditional sourdoughs. Int J Food Microbiol. 2015;215:161-70.
http://doi.org/10.1016/j.ijfoodmicro.2015.09.015

33. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11(12):2639-43.
http://doi.org/10.1038/ismej.2017.119

34. Knight R, Vrbanac A, Taylor BC, et al. Best practices for analysing microbiomes. Nat Rev Microbiol. 2018;16(7):410-22.
http://doi.org/10.1038/s41579-018-0029-9

35. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.
http://doi.org/10.1093/nar/gks1219

36. Kuczynski J, Stombaugh J, Walters WA, Gonzalez A, Caporaso JG, Knight R. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Curr Protoc Microbiol. 2012;Chapter 1:Unit 1E 5.
http://doi.org/10.1002/9780471729259.mc01e05s27

37. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC Bioinform. 2011;12:385.
http://doi.org/10.1186/1471-2105-12-385

38. Maillet A, Bouju-Albert A, Roblin S, et al. Impact of DNA extraction and sampling methods on bacterial communities monitored by 16S rDNA metabarcoding in cold-smoked salmon and processing plant surfaces. Food Microbiol. 2021;95:103705.
http://doi.org/10.1016/j.fm.2020.103705

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

39. Kim BR, Shin J, Guevarra R, et al. Deciphering diversity indices for a better understanding of microbial communities. J Microbiol Biotechnol. 2017;27(12):2089-93.
http://doi.org/10.4014/jmb.1709.09027

40. Keisam S, Romi W, Ahmed G, Jeyaram K. Quantifying the biases in metagenome mining for realistic assessment of microbial ecology of naturally fermented foods. Sci Rep. 2016;6:34155.
http://doi.org/10.1038/srep34155

41. Ercolini D, Pontonio E, De Filippis F, et al. Microbial ecology dynamics during rye and wheat sourdough preparation. Appl Environ Microbiol. 2013;79(24):7827-36.
http://doi.org/10.1128/AEM.02955-13

42. Zheng J, Wittouck S, Salvetti E, et al. A taxonomic note on the genus *Lactobacillus*: description of 23 novel genera, emended description of the genus *Lactobacillus beijerinck* 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. Int J Syst Evol Microbiol. 2020;70(4):2782-858.
http://doi.org/10.1099/ijsem.0.004107

43. Ventimiglia G, Alfonzo A, Galluzzo P, et al. Codominance of *Lactobacillus plantarum* and obligate heterofermentative lactic acid bacteria during sourdough fermentation. Food Microbiol. 2015;51:57-68.
http://doi.org/10.1016/j.fm.2015.04.011

44. Dimidi E, Cox SR, Rossi M, Whelan K. Fermented foods: definitions and characteristics, impact on the gut microbiota and effects on gastrointestinal health and disease. Nutrients. 2019;11(8): 1806.
http://doi.org/10.3390/nu11081806

45. Falasconi I, Fontana A, Patrone V, et al. Genome-assisted characterization of *Lactobacillus fermentum*, *Weissella cibaria*, and *Weissella confusa* strains isolated from sorghum as starters for sourdough fermentation. Microorganisms. 2020;8(9):1388.
http://doi.org/10.3390/microorganisms8091388

46. Hadaegh H, Seyyedain Ardabili SM, Tajabadi Ebrahimi M, Chamani M, Azizi Nezhad R. The impact of different lactic acid bacteria sourdoughs on the quality characteristics of toast bread. J Food Qual. 2017;2017:1-11.
http://doi.org/10.1155/2017/7825203

47. Fazeli MR, Shahverdi AR, Sedaghat B, Jamalifar H, Samadi N. Sourdough-isolated *Lactobacillus fermentum* as a potent anti-mould preservative of a traditional Iranian bread. Eur Food Res Technol. 2004;218(6):554-6.
http://doi.org/10.1007/s00217-004-0898-1

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

48. Litwinek D, Boreczek J, Gambus H, Buksa K, Berski W, Kowalczyk M. Developing lactic acid bacteria starter cultures for wholemeal rye flour bread with improved functionality, nutritional value, taste, appearance and safety. PLoS One. 2022;17(1):e0261677.
http://doi.org/10.1371/journal.pone.0261677

49. Yu J, Mo L, Pan L, et al. Bacterial microbiota and metabolic character of traditional sour cream and butter in Buryatia, Russia. Front Microbiol. 2018;9:2496.
http://doi.org/10.3389/fmicb.2018.02496

50. Seo DW, Yum SJ, Lee HR, Kim SM, Jeong HG. Microbiota analysis and microbiological hazard assessment in Chinese chive (*Allium tuberosum* Rottler) depending on retail types. J Microbiol Biotechnol. 2022;32(2):195-204.
http://doi.org/10.4014/jmb.2112.12013

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

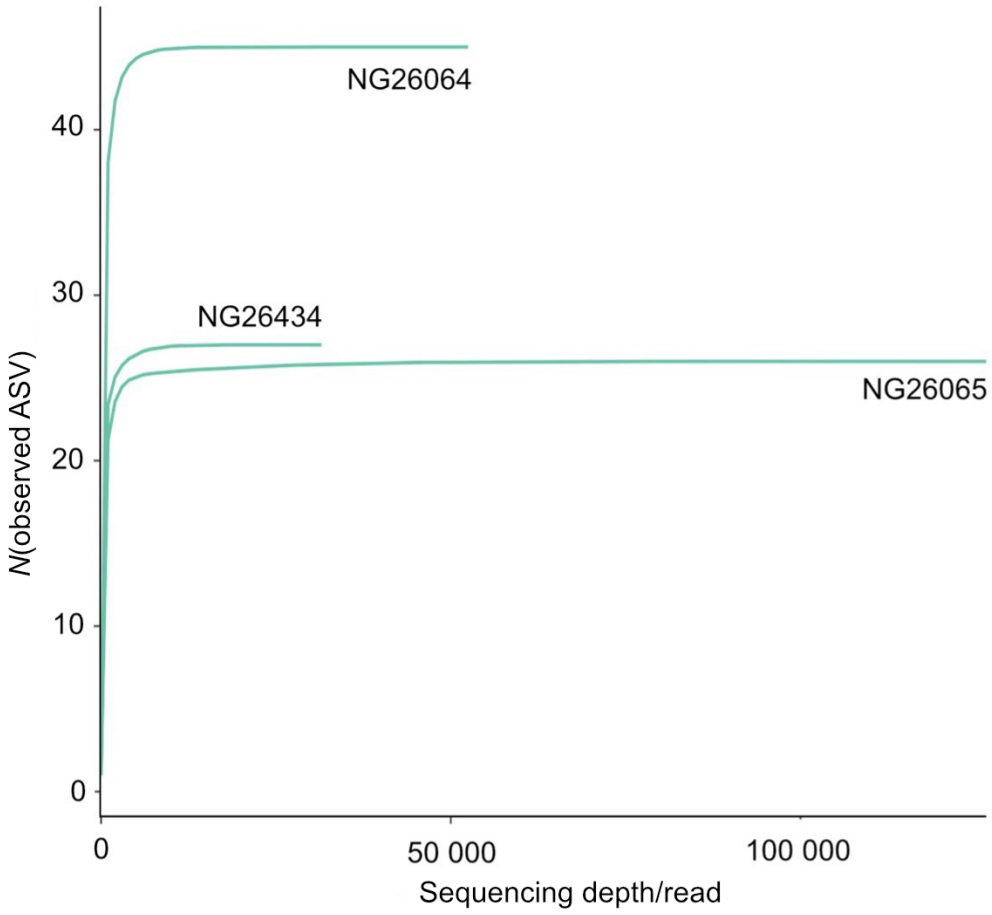

Fig. 1. Rarefaction curves constructed by using aplicon sequence variants (ASVs) observed in the three DNA isolates (NG-26064, NG-26065, NG 26434). On the abscissa are plotted the reads sequenced from DNA extracted with NucleoSpin Food kit (NG-26064), QIAamp DNA Microbiome Kit (NG-26065), and GENESpin Food kit (NG-26434). On the ordinate is shown the number of found ASVs. Each curve represents a particular isolate

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

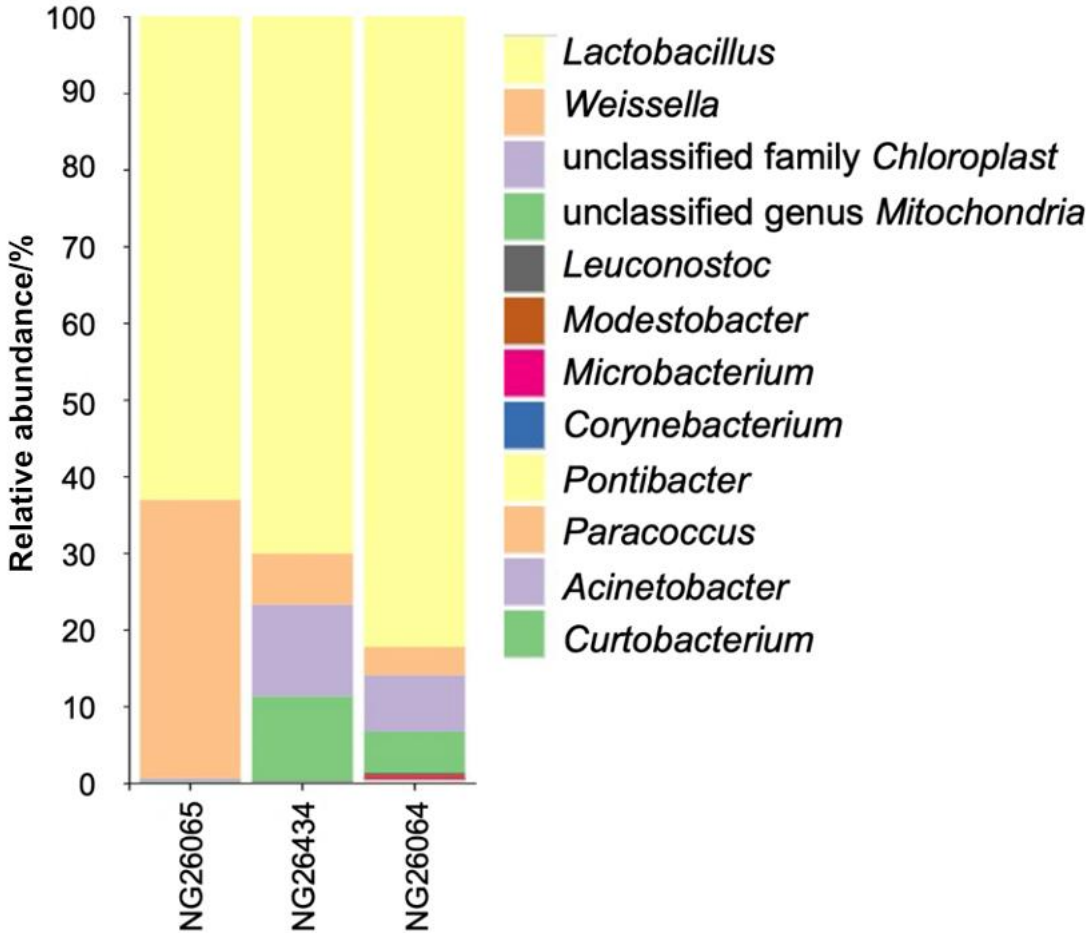

Fig. 2. Relative abundance and taxonomic assignment of bacterial community at the genus taxonomic level of the three sourdough isolates. The color key on the right shows identified genera

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

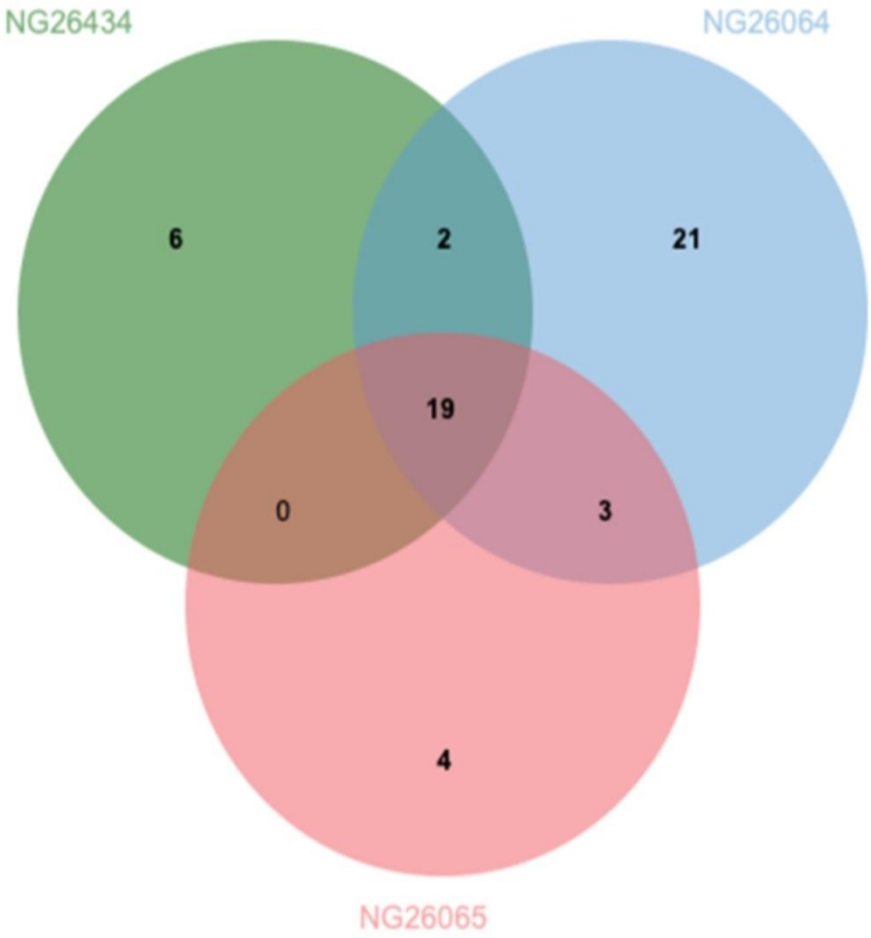

Fig. 3. Venn diagram analysis of the common and unique ASVs obtained by DNA extracted with different kits. Blue coloured area depicts ASVs found in isolate named NG-26064 obtained from NucleoSpin Food Kit (method 1), red area depicts ASVs found in isolate named NG-26065 obtained from QIAamp DNA Microbiome kit (method 2), and green colored area represents ASVs found in isolate named NG-26434 obtained from GENESpin Food Kit (method 3)

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Table 1. Main characteristics of the DNA extraction kits used

DNA extraction kit	Removal of background DNA	Cell disruption			DNA purification	*V*(elution buffer)/µL	Cost per sample/ €
		thermal	chemical	mechanical			
NucleoSpin Food Kit (Macherey - Nagel, Germany)	no	65 °C	yes	no	NucleoSpin Food Column	100 Buffer CE	4.02
QIAamp DNA Microbiome Kit (Qiagen, Germany)	yes	56 °C	yes	yes TissueLyser	QIAampUCP Mini Column	30 Buffer AVE	11.92
GENESpin Food Kit (Eurofins Technologies, Hungary)	no	65 °C	yes	no	GENE*Spin* Column	100 GENE*Spin* Buffer	2.98

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Table 2. Summary of concentration and yield of DNA extracted from sourdough with three commercial kits

Extraction method	Name of the isolate	*m*(sample)/mg	γDNA/(ng/μL)	$Y(\mathrm{DNA})_{\mathrm{total}}$/(μg)
NucleoSpin Food Kit (Macherey-Nagel, Germany)	NG-26064	200	246.8	24.68
QIAamp DNA Microbiome Kit (Qiagen, Germany)	NG-26065	2000 (2x1000)	29.2	0.88
GENESpin Food Kit (Eurofins Technologies, Hungary)	NG-26434	200	210.1	21.01

Table 3. Alpha diversity estimation of the 16S rRNA gene libraries of the sourdough from MiSeq sequencing analysis

Name of the Isolate	Reads	ASVs	ACE	Chao	Shannon	Simpson
NG-26064	52474	45	45	45	4.26	0.92
NG-26065	126553	26	26	26	2.70	0.77
NG-26434	31528	27	27	27	3.78	0.91

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

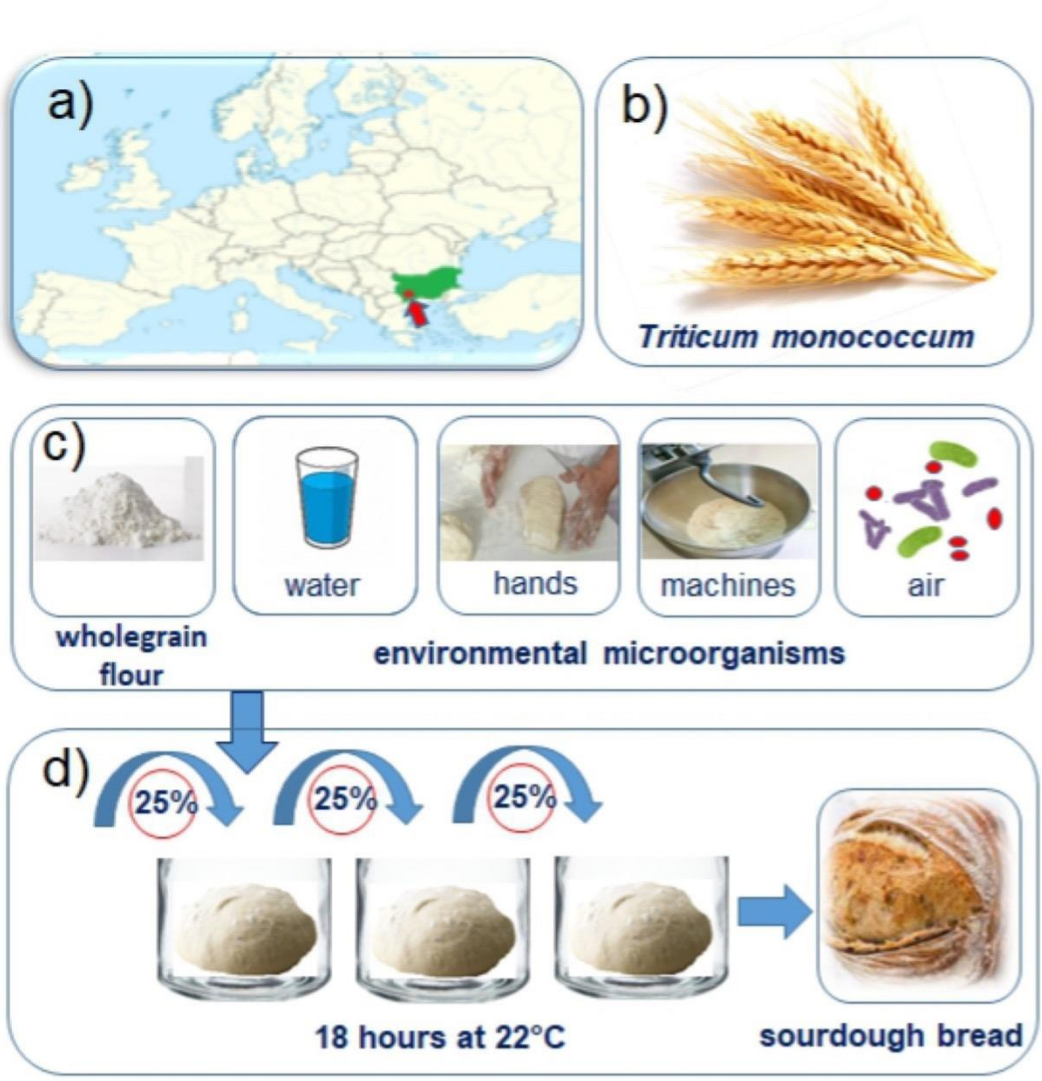

Fig. S1. Schematic representation of geographical, plant origin, and production workflow of the investigated Bulgarian sourdough. a) The sourdough was made by a manufacturer located in Southwestern Bulgaria (town of Bansko, red point). b) The type of flour used was *Triticum monococcum*. c) Ingredients of the sourdough and sources of microorganisms. d) Overview of the production workflow process. Fermentation of the sourdough proceeded 18 hours at 22 °C and included three backsloppings. For each backslopping, 1.5 % NaCl was added and 25 % of the sourdough was used