



**HAL**  
open science

## Microbial community dispersal from wheat grains to sourdoughs : a contribution of participatory research

Lucas von Gastrow, Elisa Michel, Judith Legrand, Rémy Amelot, Diego Segond, Stéphane Guezenec, Olivier Rué, Véronique Chable, Isabelle Goldringer, Xavier Dousset, et al.

### ► To cite this version:

Lucas von Gastrow, Elisa Michel, Judith Legrand, Rémy Amelot, Diego Segond, et al.. Microbial community dispersal from wheat grains to sourdoughs : a contribution of participatory research. *Molecular Ecology*, 2022, 32 (10), pp.2413-2427. 10.1111/mec.16630 . hal-03741463v1

**HAL Id: hal-03741463**

**<https://hal.inrae.fr/hal-03741463v1>**

Submitted on 1 Aug 2022 (v1), last revised 7 Nov 2023 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Microbial community dispersal from wheat grains to sourdoughs : a contribution of participatory research

Lucas von Gastrow<sup>a,b,\*</sup>, Elisa Michel<sup>a,c,1</sup>, Judith Legrand<sup>d</sup>, Rémy Amelot<sup>a</sup>, Diego Segond<sup>a</sup>, Stéphane Guezenec<sup>a</sup>, Olivier Rué<sup>e,f</sup>, Véronique Chable<sup>g</sup>, Isabelle Goldringer<sup>d</sup>, Xavier Dousset<sup>c</sup>, Estelle Serpolay-Besson<sup>h</sup>, Bruno Taupier-Letage<sup>h</sup>, Camille Vindras-Fouillet<sup>h</sup>, Bernard Onno<sup>c</sup>, Florence Valence<sup>b</sup>, and Delphine Sicard<sup>a</sup>

<sup>a</sup>SPO, INRAE, Montpellier SupAgro, Montpellier 34000, France

<sup>b</sup>STLO, INRAE, Institut Agro, Rennes Cedex 35042, France

<sup>c</sup>Oniris, Laboratoire MicrobioTech, UMR GEPEA 6144, Rue de la Géraudière CS 82225, 44322 Nantes Cedex 3, France

<sup>d</sup>Génétique Quantitative et Evolution le Moulon, Université Paris-Sud, INRAE, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France

<sup>e</sup>Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France

<sup>f</sup>Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

<sup>g</sup>INRAE, UMR BAGAP, 65 rue de Saint Briec, 35042 Rennes, France

<sup>h</sup>Institut Technique de l'agriculture et de l'Alimentation Biologique, 149 Rue de Bercy, 75012 Paris, France

<sup>1</sup>should be considered joint first authors

\*Corresponding author : Lucas von Gastrow, lucas.gastrow@gmail.com

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1111/1.1.0.1](https://doi.org/10.1111/1.1.0.1)

Accepted Article

## 1 Abstract

2 Understanding microbial dispersal is critical to understand the dynamics and evolution of microbial  
3 communities. However, microbial dispersal is difficult to study because of uncertainty about their  
4 vectors of migration. This applies to both microbial communities in natural and human-associated  
5 environments. Here, we studied microbial dispersal along the sourdoughs bread making chain  
6 using a participatory research approach. Sourdough is a naturally fermented mixture of flour and  
7 water. It hosts a community of bacteria and yeasts whose origins are only partially known. We  
8 analysed the potential of wheat grains and flour to serve as an inoculum for sourdough microbial  
9 communities using 16S rDNA and ITS1 metabarcoding. First, in an experiment involving farmers,  
10 a miller and bakers, we followed the microbiota from grains to newly initiated and propagated  
11 sourdoughs. Second, we compared the microbiota of 46 sourdough samples collected everywhere  
12 in France, and of the flour used for their backslopping. The core microbiota detected on the seeds,  
13 in the flour and in the sourdough was composed mainly of microbes known to be associated with  
14 plants and not living in sourdoughs. No sourdough yeast species were detected on grains and flours.  
15 Sourdough lactic acid bacteria were rarely found in flour. When they were, they did not have the  
16 same amplicon sequence variant (ASV) as found in the corresponding sourdough. However, the  
17 low sequencing depth for bacteria in flour did not allow us to draw definitive conclusion. Thus,  
18 our results showed that sourdough yeasts did not come from flour, and suggest that neither do  
19 sourdough LAB.

## 20 Keywords

21 Microbial ecology, dispersal, yeast, lactic acid bacteria, bread, fermentation

# 1 Introduction

Understanding the functioning and evolution of communities is central to ecological studies. Many of the concepts and debates that have animated this field have arisen from the study of plant communities (Mikkelsen, 2005). Microbial communities have also been a subject of increasing interest, and it is now clearly established that they play a central role in the functioning and evolution of many ecosystems. Numerous concepts have been proposed in community ecology but it is only recently that theoretical models have unified them to take account of local evolutionary dynamics and the links between communities. Vellend (2010) defined four factors that shape communities: diversification, selection, dispersal and drift. More recently, Thompson et al. (2019) proposed a meta-community model with three factors : density-independent responses to abiotic conditions, density-dependent biotic interactions, and dispersal. These general frameworks offer valuable tools to understand the dynamics of microbial communities but suffer from a lack of empirical data on the selection processes and dispersal of microbial communities.

Microbial dispersal has been mainly studied in natural environments or for human pathogen. The dispersal of beneficial microorganisms in agri-food chains has received less attention. In agri-food chains, humans are expected to contribute to the dispersal of microorganisms, either as a direct vector of transmission or indirectly through the food products they carry. A high dispersal rate would lead to a reduction of diversity throughout the food chain and between food chains. Dispersion through food products could have even greater consequences for long food supply chain, where products are transported over long distance. It is therefore important to better understand the dispersion of microorganisms in food chains to better predict their ability to maintain microbial biodiversity and therefore, their sustainability.

Microbial communities have been used to make fermented foods since the Neolithic era (Tamang and Kailasapathy, 2010) in which they usually display relatively little complexity with regards to wild environments, making them good model systems for ecological studies. They are organized as

47 metacommunities in which the microbial community of each leaven evolves as a function of human  
48 practices and may be linked to others through exchanges of the fermented product themselves or  
49 of the raw materials used to make them. Among the numerous fermented foods, sourdough mi-  
50 crobial communities used for bread-making represent a good metacommunity model system. First,  
51 sourdough microbial communities are relatively simple, usually containing one to two dominant  
52 bacterial and yeast species (Carbonetto et al., 2018; Arora et al., 2021). Second, sourdoughs  
53 are made of few ingredients, basically flour and water, which are regularly added to feed the  
54 microorganisms, thus limiting the number of sources of microbial species. Third, the microbial  
55 communities in sourdough have been widely studied and reviewed (De Vuyst et al., 2016; Gänzle  
56 and Ripari, 2016; Gobbetti et al., 2016; Gänzle and Zheng, 2019; Arora et al., 2021; Van Ker-  
57 rebroeck et al., 2017; Landis et al., 2021; Calvert et al., 2021; Lau et al., 2021). Well known  
58 species such as *Fructilactobacillus sanfranciscensis*, *Lactiplantibacillus plantarum*, *Levilactobacil-*  
59 *lus brevis* bacteria and *Saccharomyces cerevisiae*, *Kazachstania humilis*, *Torulaspota delbrueckii*  
60 and *Wickerhamomyces anomalus* yeasts are frequently encountered. Finally, population genomic  
61 analysis of the yeast species *S. cerevisiae* has shown that sourdough yeast populations differ  
62 from commercial yeasts and may have undergone specific selection processes when compared to  
63 industrial processes (Bigey et al., 2020). The diversity of microbial communities has been stud-  
64 ied in home-made and bakery sourdoughs from all over the world (Landis et al., 2021; Arora  
65 et al., 2021) and did not display any clear geographical pattern. Several studies revealed that  
66 the bakery house microbiota is the main driver of sourdough microbial diversity but the origin of  
67 microbes that composed the house microbiota remained to be elucidated. The same species of  
68 lactic acid bacteria (LAB) or yeast could be found on the baker's tools (Minervini et al., 2015)  
69 or hands (Reese et al., 2020) and in their sourdough. But, no sourdough microorganisms were  
70 detected in the bakery air (Minervini et al., 2015) or in the water (Scheirlinck et al., 2009; Reese  
71 et al., 2020) used to make the sourdough. Some studies have shown that flour can be a vector for  
72 *Lactobacillaceae*. However, this was only shown for three different flours (Minervini et al., 2018a)

73 or for laboratory-made sourdoughs (De Angelis et al., 2019), whose dynamics are not the same  
74 as bakery sourdoughs (Minervini et al., 2012). It is still unclear whether flour is a source or/and  
75 a vector of sourdough yeast and bacteria.

76 In this paper, we tested whether wheat grains and flour can be a source of microorganism  
77 in newly initiated sourdough and in mature sourdoughs. The study was carried out in French  
78 sourdoughs which was found to include most of the yeast species detected world-wide (Carbonetto  
79 et al., 2018) and the main lactic acid bacteria species (Lhomme et al., 2015; Michel et al., 2016).  
80 The distribution of microbial species in this country suggests that flour could be a vector of  
81 dispersion. The majority of sourdoughs contained *F. sanfrancicensis* suggesting that this species  
82 is easily dispersed among the sourdoughs, potentially through the flour (Lhomme et al., 2015;  
83 Michel et al., 2016). The diversity of yeast species appeared to be structured by the bread  
84 making process. Sourdoughs made with farmer-baker's practices mostly exhibit *Kazachstania*  
85 *bulderi*, while sourdoughs made with artisans baker's practices mostly exhibit *Kazachstania humilis*,  
86 suggesting a potential role of the type of flour used (Urien et al., 2019). We used a participatory  
87 research approach including farmers, bakers and scientists. First, we set up an experiment where  
88 6 wheat populations were used to make 6 flours, which were given to four bakers that make  
89 sourdoughs and we follow up microorganism dispersion along the sourdough making chain from  
90 grain to mature sourdough. Second, we studied the microbial species diversity of 44 flours and  
91 related mature sourdough samples collected in French bakeries. We did not find any evidence that  
92 flour is a vector for sourdough yeasts. Flour rarely carry LAB species and at very low density. We  
93 discuss these findings in relation with ecological processes driving microbial community evolution  
94 in sourdough.

## 2 Material and Methods

### 2.1 Participatory experimental design for studying microbial community succession from grain to sourdough

#### 2.1.1 Seeds sampling

Three wheat landraces, (i) 'Saint Priest et le Vernois Rouge', (ii) 'Redon Roux Pâle' and (iii) 'Bladette de Provence' and three modern wheat varieties, (i) Renan, (ii) Pireneo and (iii) Chevalier were grown each on three terroirs. The seeds were sown in autumn 2014 and harvested in July 2015. The grains were sorted with an optical sorter, an air column and a cleaner-separator. They were stored in dehydrated room in 20-25°C till the end of October, then in 10°C without dehydration to fight against weevils. For each terroir, grains from the three landraces were mixed and so were grains from the three modern varieties leading to six lots of grains. Two independent samples of 250 gr of these six lots (12 samples) as well as of grains of each varieties from two terroirs (24 samples) were collected, resulting in 26 samples that were stored at -20°C for metabarcoding analyses.

#### 2.1.2 Milling and flours sampling

The six lots of grains were sent to a French miller (Hélène Chaudy, Ferme d'Orvillier) who is also a farmer-baker (Baker B3). Crushing and sifting were performed with an Astrie stone-mill. To avoid flour contaminations from one lot to another, the miller ground in alternation one experimental lot and his own grains. For each of the six grain lot, samples of flours were collected at the beginning and at the end of the milling in duplicate leading to 24 flour samples. Flour made from the miller own seeds was also sampled at the beginning and at the end of the milling process as control. The 26 flour samples were stored at -20°C for metabarcoding analyses.

### 117 **2.1.3 Sourdough sampling**

118 The flours of the six seeds lots were sent to four bakers (MP, PV, HS, LM), who initiated and  
119 maintained six sourdoughs for three weeks. Sourdoughs were also initiated at the lab in the same  
120 conditions (LL). Each baker received a pallet with 10 kg of each flour, buckets for each sourdough  
121 and sample pots for the sampling. They also got a scoresheet to follow the temperature of the  
122 flour, of the bakery, of the water; the weight of flour and water at each backslopping. Each  
123 baker made the sourdoughs with his own protocol, although the same one had to be used for  
124 the six received flours. They all had to maintain their sourdough for three weeks. Baker MP,  
125 PV, HS, LM made 5, 7, 9, and 10 backslopping respectively. Bakers were asked to have their  
126 hands washed between each sourdough back-slopping. The bakers took samples of sourdough,  
127 approximately 40 grams, the day of the initiation, after one week, two weeks and three weeks of  
128 back-sloppings and stored the samples at -20°C for metabarcoding analysis. The same protocol  
129 of sourdough preparation and sampling was done in a laboratory environment as control although  
130 13 backslopping were carried out in 3 weeks. Overall, 128 samples of sourdoughs were stored for  
131 metabarcoding analyses including i) 30 experimental sourdoughs J0 ii) 30 sourdoughs collected  
132 at sampling time S1 (week 1), iii) 30 sourdoughs collected at sampling time S2 (week 2) and iv)  
133 30 sourdoughs collected at sampling time S3 (week 3) v) 8 baker's own sourdoughs, 4 collected  
134 at the beginning of the experiment and 4 collected at the end of the experiment.

## 135 **2.2 Collection of flour and sourdough from French bakers**

### 136 **2.2.1 Survey of bread-making practices**

137 A total of 22 bakers and 22 French farmer-bakers completed a questionnaire on their bread-making  
138 practices, as described by Michel et al. (2019). Questions concerned sourdough management  
139 (addition of bran, back-slopping technique, time elapsing since sourdough initiation, sourdough  
140 hydration, number of back-slopping procedures per week and between bread-making sessions,



141 temperature at back-slopping), the flour (self-produced or not, type of cereal variety, type of  
142 mill) and the bread-making process (use of selected baker's yeast in bread or in other products,  
143 mechanical or manual kneading, proportions of sourdough, flour, water and salt in bread dough,  
144 fermentation time, quantity of bread produced each week, number of bread-making sessions per  
145 week). We also asked the producers if they had shared raw materials (grains, flour or sourdough)  
146 or if they had physical contacts with each other. This survey, together with the survey from  
147 Michel et al. (2019) were used to choose the experimental design presented above (section 2.1).

### 148 **2.2.2 Sample collection**

149 A total of 46 mature sourdoughs were collected, together with the flour used to make each one.  
150 The collection included only sourdoughs made by bakers but no home-made sourdoughs, because  
151 the process of domestication and dispersal could be different with these two types of sourdoughs.  
152 Forty-four sourdoughs came from different bakeries located everywhere in France, and two bakeries  
153 (B64 and B68) sent two sourdoughs, so that 46 sourdough and 44 flour samples were studied.  
154 We chose to collect samples at country scale because the millers who produce the fresh flour for  
155 bakers, distribute it mainly at this geographical scale. Samples were collected between September  
156 2018 and July 2019 and were received at the laboratory within one to three days. All samples  
157 were stored at -20°C in plastic bags and plastic tubes, respectively, before DNA extraction.

### 158 **2.2.3 Sourdough and flour microbial enumeration**

159 All 46 sourdoughs and 21 of the 44 flour samples were plated at reception. 10 g sourdough or 3  
160 g flour were diluted ten times in tryptone-salt buffer (1 g/L tryptone, 8 g/L NaCl). After serial  
161 dilutions, lactic acid bacteria (LAB) were enumerated on MRS-5 (Meroth et al., 2003) with 100  
162 µg cycloheximide and on PCA (6 g/L Tryptone, 2.5 g/L yeast extract, 1 g/L glucose, 15 g/L agar)  
163 media while yeasts were enumerated on YEPD medium (10 g/L yeast extract, 20 g/L peptone,  
164 20 g/L dextrose, 100 mg/L chloramphenicol). Culture-dependent methods were only used for

165 enumeration purposes.

## 166 **2.3 Metabarcoding analysis**

167 Microbial species diversity of grains, sourdoughs and flours was analyzed by amplicon-based DNA  
168 metabarcoding. For the collected sourdoughs, two separate Illumina MiSeq runs were used for  
169 sourdough and flour to prevent any contamination between the sample types.

### 170 **2.3.1 DNA extraction from seeds, flour and sourdough**

171 DNA from seeds microbiota was obtained after maceration of ten grams of seeds in 40 mL of  
172 buffer PBS Tween 20 (0.05% v/v) for 40 hours at 4°C with constant stirring (200 rpm). The  
173 macerates were filtered with Stomacher® bags. After centrifugation at 5000 rpm, during 10 min  
174 at 4°C, 2 mL of the supernatant was mixed with the pellet and DNA was extracted using the  
175 Powersoil DNA isolation kit (MoBio) as explained in the manufacturer's protocol.

176 Flour DNA was obtained by mixing five grams of each flour with 35 mL of PBS buffer using  
177 a stomacher during 5 min. The filtrate was centrifuged at 1000 rpm, for 5 min at 4°C. The  
178 supernatant, was then centrifuged at 2500 rpm for 15 min at 4°C. The pellet was re-suspended in  
179 500 µL of PBS buffer and the DNA extracted following the Powersoil DNA isolation kit procedure  
180 (MoBio 12888-100).

181 Finally, sourdough DNA was extracted directly from 200 mg of material following the Qiagen  
182 PowerSoil DNA isolation kit procedure (12888-100). Extraction was then performed in accordance  
183 with the manufacturer's instructions.

### 184 **2.3.2 Markers amplification and MiSeq sequencing**

185 The 16S V3-V4 region was amplified for bacteria and the ITS1 region for fungi. For fungi, the  
186 ITS1 region was targeted with the PCR primers ITS1-F (5' - CTTGGTCATTTAGAGGAAGTAA  
187 - 3') and ITS2 (5' - GCTGCGTTCTTCATCGATGC - 3') (White et al., 1990), while for bacteria,

188 the 16V3-V4 region was targeted with the PCR primers 343F : (5' - TACGGRAGGCAGCAG -  
189 3') and 784R : (5' - TACCAGGGTATCTAATCCT - 3') (Liu et al., 2007). The primers also  
190 included the Illumina tail (5' - TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG - 3'), and a  
191 frame-shift of four, six or eight random nucleotides for forward primers and four, five or six  
192 random nucleotides for reverse primers, in order to prevent saturation during sequencing. The  
193 resulting primers therefore had the following structure: 5' - Illumina tail - frame-shift - genome  
194 targeting region - 3'. All the primers used are listed in Supplementary Materials Table S1. For each  
195 forward or reverse primer, an equimolar mix of the three primers containing the different frame-  
196 shifts was added to the PCR mix. To prepare the multiplexed Illumina libraries, we employed  
197 a strategy based on a two-step PCR approach : a first PCR using the locus-specific primers  
198 including the Illumina adapter overhang (with 30 cycles), and a second PCR for the incorporation  
199 of Illumina dual-indexed adapters (with 12 cycles). Bead purifications were carried out after each  
200 PCR. Quantification, normalization and pooling were performed before sequencing on Illumina  
201 MiSeq (Ravi et al., 2018).

### 202 2.3.3 Bioinformatics analyses

203 The resulting sequences were analyzed using R (Team, 2019) workflow combining dada2 v.1.16 (Calla-  
204 han et al., 2016) and FROGS 3.2.2 (Escudié et al., 2018) software. Reads were filtered, merged  
205 and assigned to ASVs with dada2 and the ASVs were assigned to species using the FROGS affil-  
206 iation tool. Adapters were first removed using cutadapt v. 1.12 with Python 2.7.13. Reads were  
207 then filtered using the dada2 filterAndTrim function, with a truncation length of 250 bp for ITS1  
208 forward and reverse reads and 275 and 200 bp for 16S forward and reverse reads, respectively. This  
209 truncation reduced the error rate while still allowing the merging of most reads. The error model  
210 was then calculated using the learnErrors function. Reads were dereplicated using derepFastq and  
211 the dada2 core sample inference algorithm was executed. Forward and reverse reads were then  
212 merged with a minimum overlap of 20 bp. The resulting sequences were saved in a sequence ta-

213 ble using makeSequenceTable. Chimera were removed using the removeBimeraDenovo function.  
214 The amplicon sequence variants (ASV) in the sequence table were then assigned to species using  
215 FROGS affiliation v3.2.2 with silva 138 (Quast et al., 2013) for 16S and Unite 8.0 (Nilsson et al.,  
216 2019) for ITS1. Unite was completed with ITS1 reference sequences from yeast species usually  
217 found in sourdough. Multi-affiliations were dealt with by assigning the lowest common taxonomy  
218 level to multi-affiliated ASVs. Samples were rarefied to the minimum number of reads for each  
219 barcode, or 1000 reads using the rarefy\_even\_depth function of the R (v. 4.1.0) phyloseq pack-  
220 age (v. 1.24.2) (McMurdie and Holmes, 2013). Samples with a depth of less than 1000 were  
221 discarded. If not otherwise specified, the analyses were conducted on the rarefied data.

#### 222 **2.3.4 Analysis of bread-making practices**

223 Groups of bread-making practices were obtained with an MCA computed with the R package  
224 FactoMineR (v. 2.4), and individuals were clustered using the HCPC function with two clusters.  
225 They were plotted using the factoextra package (v. 1.0.7).

#### 226 **2.3.5 Statistical analysis**

227 A Wilcoxon-Mann-Witney test was performed to compare the diversity index between the flour  
228 and sourdough samples. The correlation between flour and sourdough diversity was computed  
229 using a Spearman rank-order correlation test. Both tests were computed using the R package stats  
230 v 3.6.2, with the wilcox.test and cor.test functions, respectively. A Mantel test was performed to  
231 test the link between geographical distances for sourdoughs and Bray-Curtis distance matrices,  
232 using the R ape package (v. 5.5) mantel.test function.

## 233 3 Results

### 234 3.1 Microbial community succession along the sourdough making 235 chain

236 To investigate the potential role of wheat grains and flour as an inoculum of yeast and bacteria in  
237 newly initiated sourdoughs, we investigate the composition of the microbiota along the sourdough  
238 making chain following the sequence of wheat seeds, flours, sourdoughs at initiation, and after one,  
239 two and three weeks of backslopping. Overall, a total of 190 samples, including 36 wheat grains,  
240 24 flours and 120 sourdough samples were analysed by metabarcoding on the 16S V3V4 region  
241 for bacteria and ITS1 gene for fungi. After trimming and removal of sequences corresponding to  
242 mitochondria and chloroplasts, we obtained a median of 38013, 2178 and 37165 reads for grains,  
243 flours and sourdoughs samples respectively for 16S sequencing, and 20818, 19597 and 19125 for  
244 ITS1 sequencing (Supplementary figure S1).

245 The level of alpha diversity on the seeds and in the flours were close with 21 and 16 bacterial  
246 genus and 33 and 37 fungal genus detected on grains and in flour respectively. Species richness  
247 ranged from 1 to 40 per grain sample and 4 to 18 per flour sample for bacteria and from 6 to  
248 38 and 8 to 29 in grain and flour samples respectively for fungi (Figure 1). The level of alpha  
249 diversity then decreased over the weeks as sourdoughs were back-slopped to reach an average  
250 of 13 bacteria and 13 yeast species in three weeks old sourdoughs (Figure 1). Three weeks  
251 old sourdoughs were the only samples to have dominant genus (Freq >0.5). Their microbial  
252 communities were indeed typical of a mature sourdough, with one to three dominant species of  
253 bacteria and yeasts except for laboratory sourdoughs and sourdoughs made by baker 1 that had  
254 only carried five backslopping during the three weeks. Surprisingly, a lower bacteria alpha diversity  
255 was observed in flour compared to newly initiated sourdough. This observation could be explained  
256 by the low number of reads in flour samples after having removed the sequences corresponding  
257 to chloroplasts and mitochondria (Supplementary figure S1).

258 The core microbiota, which can be defined as the microbial community consistently present  
259 in grains, flours, and sourdoughs, was composed of 10 genera of bacteria with a high prevalence  
260 of *Pantoea* and *Pseudomonas*. Other genera included *Erwinia*, *Massilia*, *Paenibacillus*, *Sphin-*  
261 *gomonas*. These genera were strongly represented on the grains, but their frequency decreased  
262 along the sourdough making chain. The composition of the bacterial genera on the grains and in  
263 the flours was similar, with the exception of one batch of flour containing mainly the genus *Erwinia*.  
264 In contrast, several bacterial genera were detected only in the sourdoughs, including *Weisseila*, as  
265 well as several genera of the family *Lactobacillaceae*. The proportion of the latter increased until  
266 it composed almost all the reads after three weeks of backslopping. Different species, includ-  
267 ing *Fructilactobacillus sanfranciscensis*, *Companilactobacillus paralimentarius*, *Lactiplantibacillus*  
268 *plantarum*, *Levilactobacillus brevis* increased in frequency depending on the baker.

269 Concerning fungi, the core microbiota was composed of 26 genus, with a prevalence of *Al-*  
270 *ternaria*, *Epicoccum*, *Mycosphaerella*, *Cladosporium*, *Filobasidium*, *Cystofilobasidium* and *Vishni-*  
271 *acozyma*. The genera *Fusarium* and *Penicillium*, known to produce mycotoxin, were also present  
272 all along the chain but with a lower frequency. The frequency of fungal phylum (Ascomycota  
273 and Basidiomycota) changed from the grains to three weeks old sourdoughs, with Basidiomycota  
274 mostly found in seeds (freq >0.6) and Ascomycota mostly found in sourdoughs S3 (freq >0.8).  
275 The fermenting yeast genera *Saccharomyces*, *Kazachstania*, *Torulasporea*, and *Wickerhamomyces*  
276 appeared during backsloppings and were dominant after three weeks in the sourdoughs. Within a  
277 bakery, the same species became dominant regardless of the flour used for backslopping. As for  
278 bacteria, different yeast species invaded the community depending on the bakery suggesting an  
279 origin from the house microbiota. They included *Saccharomyces cerevisiae*, *Kazachstania bulderi*,  
280 *Kazachstania servazzii*, *Wickerhamomyces anomalus*.

281 Overall, the *Lactobacillaceae* and *Saccharomycetaceae* that are generally present in sourdough  
282 and known to be responsible for most of its beneficial functions were detected after sourdough  
283 initiation but were not found in the grains and flours, except in two flour samples. One flour

284 sample contained 0.01 percent of *F. sanfranciscensis* and another 0.06 percent of *Kazachstania*  
285 *exigua*. However, the ASVs found in the flour were never found in the sourdough made with this  
286 flour.

### 287 **3.2 Mature sourdoughs microbiota had greater microbial density but** 288 **less species diversity than the flour microbiota**

289 We also compared 46 mature sourdough samples obtained from 44 bakeries located throughout  
290 France with the 44 flour samples used to refresh them, in order to confirm on a larger scale that  
291 flours do not contain fermentative species found in sourdough.

292 On average, microbial density was higher in sourdoughs than in flours, for both bacteria and  
293 fungi. Sourdoughs contained on average  $1.9 \times 10^7$  (sd =  $1.3 \times 10^7$ ) CFU/g (colony forming  
294 units/g) of yeast while flours contained a mean of  $2.3 \times 10^3$  (sd =  $1.6 \times 10^3$ ) CFU/g. As for  
295 bacteria, the sourdoughs contained  $1.3 \times 10^9$  (sd =  $1.3 \times 10^9$ ) CFU/g while flours contained  
296  $7.7 \times 10^3$  (sd =  $2.0 \times 10^4$ ) CFU/g or  $6.9 \times 10^4$  (sd =  $1.0 \times 10^5$ ) CFU/g, depending on whether  
297 the estimation of bacterial density was performed on MRS or PCA. Sourdoughs were only plated  
298 on MRS medium, as we expected to find only *Lactobacillaceae*, while flour generally harbors a  
299 more diverse bacterial community so we also plated these samples on PCA, which is a less specific  
300 medium. The observation of fungal morphology on YEPD petri dishes revealed that most flour  
301 samples contained filamentous fungi, some with a typical *Penicillium* morphology, while sourdough  
302 samples were characterized by the presence of yeasts.

303 The composition of the flour and sourdough microbiota was then analyzed by metabarcoding  
304 on the 16SV3V4 and ITS1 regions. After trimming and removal of sequences corresponding to  
305 mitochondria and chloroplasts, we obtained a median of 2031 and 62421 reads for flours and  
306 sourdoughs respectively for 16S sequencing, and 27371 and 44567 reads for ITS1 sequencing.  
307 Although sourdoughs had a higher microbial density than flour, their microbial communities were  
308 less diverse than those in flour. Alpha diversity indexes calculated on the number of bacterial

309 and fungal species were significantly lower in sourdough than in flour in terms of both richness  
310 (Wilcoxon-Mann-Witney test, bacteria  $W = 1725.5$ ,  $P < 0.001$ , fungi  $W = 1555.5$ ,  $P < 0.001$ )  
311 and evenness (Wilcoxon-Mann-Witney test, bacteria  $W = 1929$ ,  $P < 0.001$ , fungi  $W = 1467$ ,  
312  $P < 0.001$ ; Figure 3). This difference was greater for bacteria than for fungi, with averages of  
313 four and 11 species for bacteria in sourdough and flour, respectively, and 10 and 13 species for  
314 fungi in sourdough and flour, respectively.

315 Sourdough species diversity was not correlated with flour species diversity for either bacteria  
316 (Spearman = 13617,  $P = 0.86$ ) or fungi (Spearman = 13019,  $P = 0.91$ ).

317 The microbiota compositions of sourdough and flour were characterized by different families.  
318 The bacteria in the sourdoughs were almost entirely composed of *Lactobacillaceae*, while flour  
319 contained mainly *Erwiniaceae* and *Pseudomonadaceae*. In sourdough, all samples but three con-  
320 tained *Fructilactobacillus sanfranciscensis* as the dominant bacterial species; the others contained  
321 *Companilactobacillus paralimentarius*. Less frequently, the presence of *Levilactobacillus brevis*,  
322 *Latilactobacillus* sp. and *Lactilactobacillus* sp. was found. In flour, *Erwiniaceae*, *Pantoea ag-*  
323 *glomerans*, an unidentified *Pantoea* sp., and *Pseudomonadaceae* were generally detected. Among  
324 *Pseudomonas* sp., some were *P. graminis*, *P. rhizosphaerae* or *P. donghuensis*. As for fungi, *Sac-*  
325 *charomycetaceae* was determined in most sourdough samples but was almost absent from flour  
326 samples (Figure 4); *S. cerevisiae* was found in 14 sourdough samples, *K. humilis* in seven samples  
327 and *K. bulderi* in six. These species were never found in flours. *Pleosporaceae* species (*Al-*  
328 *ternaria alternata* and *Alternaria infectoria*), *Mycosphaerellaceae* (*Mycosphaerella tassiana*) and  
329 an unidentified fungus from the *Dothideomycetes* family were detected at a high frequency in  
330 almost all flour samples.



### 3.3 Very little overlap between the microbiotas of mature sourdough and flour

Any overlaps between the mature sourdough and flour communities were analyzed using the Weighted Bray-Curtis distance calculated on the basis of species diversity. The Weighted Bray-Curtis was used to build two PCoAs, one for the bacterial community and the other for the fungal community. PCoA axis 1 and 2 explained 79.1% and 8.5% of variance for bacteria, and 28.5% and 13.6% of variance for fungi (Figure 5). For bacteria, axis 1 separated the flour and sourdough bacterial communities. For fungi, axis 1 separated many but not all of sourdough fungal communities from flour communities. Over the 46 sourdough fungal communities, 14 co-localized with flour fungal communities. Flour and sourdough dissimilarity matrices were not correlated (Mantel test,  $z = 836$ ,  $p = 0.667$  for bacteria and  $z = 854$ ,  $p = 0.13$  for fungi). Close microbial communities among flours did not lead to close microbial communities among sourdoughs.

We analyzed bread-making practices in order to determine whether they might be related to microbial communities in sourdough and flour. Two groups of bread-making practices could be distinguished. Farmer-baker practices (cluster 1) were more frequently associated with the use of non-commercial yeast, ancient wheat landraces, small production runs and lengthy fermentation while artisanal practices (cluster 2) were generally characterized by larger scale production, short fermentation, and the use of commercial yeast and modern wheat varieties. Sourdough from farmer-bakers frequently contained *K. bulderi* as the dominant yeast species. However, analysis of the association between sourdough and flour microbial community dissimilarity and the geographical distances between bread-making practices did not reveal any correlation (Mantel test, for flour,  $z = 308$ ,  $p = 0.59$  and  $z = 235$ ,  $p = 0.79$  for bacteria and fungi, respectively; for sourdough,  $z = 153$ ,  $p = 0.60$  and  $z = 411$ ,  $p = 0.32$  for bacteria and fungi, respectively).

The differences between the microbial communities in sourdough and flour were explained by the high abundance in sourdough samples of fermentative microorganisms, which were almost

356 never found in the flour samples. (Figure 6).

357 Overall, fermentative bacteria in the *Lactobacillales* order and yeast in the *Saccharomycetales*  
358 order were not detected in most flour samples. Out of 46 samples, ten flour samples contained  
359 fermentative bacterial species (*F. sanfranciscensis*, *Lactococcus garviae*, *Carnobacterium diver-*  
360 *gens*, *Weisella* or *Streptococcus* species) and 13 harbored fermentative yeasts (*Candida saitoana*,  
361 an unidentified *Candida* species, *Wickerhamomyces anomalus*, *Mechnikovia* sp. or *Eremothecium*  
362 *coryli*). However, the fermentative species found in flour samples were generally not found in  
363 the related sourdoughs. In six cases, *F. sanfranciscensis* was found in both flour and sourdough.  
364 Nevertheless, in these cases, the ASVs were not the same except in the case of baker 53 (Fig-  
365 ure 7). *Lactococcus garviae* was found in the flour and sourdough used by baker 45 but only  
366 one read was present in the sourdough and this ASV differed from that found in the flour. An  
367 unidentified *Metschnikovia* species was found in four pairs of sourdough and flour, and *Candida*  
368 *saitoana* and an unidentified *Candida* species in one pair of sourdough and flour samples, although  
369 the same ASV was not found in them. Many non-fermentative fungal species were shared between  
370 flour and sourdough samples. They were mainly filamentous fungi, and notably species from the  
371 genus *Alternaria* or *Mycosphaerella*. For these species, the flour and sourdough samples shared  
372 on average 0.98 ASV (sd = 1.48).

## 4 Discussion

### 4.1 Interests and limits of the participatory approach

This study could not have been conducted without developing a participatory approach. Farmers, Bakers and scientists all engaged during a four years project to gain new insight into the microbial diversity of the bakery food chain with the aim of promoting the sustainability of this food chain. Farmers and bakers contributed by designing the experiments, growing wheats, milling grains and making experimental sourdoughs. The choice of wheat populations (two types of grain mixture representing landraces and modern varieties), milling technique (stone-mill), sourdough elaboration practices (hydration, number of week of backslopping, ...) were discussed by professionals and researchers. Bakers and farmer-bakers also sent samples of their sourdough and flour for analysis at the laboratory. In addition, they participate to the interpretation of the data and dissemination of the results. Overall, 54 farmers, farmer-bakers and bakers participated to the study. This approach allowed us to analyze the dispersion of microbes along the sourdough making chain in real agri-food environments. Laboratory-made sourdoughs typically have a different microbial composition than those made in bakeries as show by our study (Figure 2) and others such as in Minervini et al. (2012). Therefore, the analysis of sourdough ecology requires the study of sourdough *in situ*, which requires the active participation of bakers. This collaboration also allows for knowledge sharing, with bakers learning microbial ecology and scientists learning baking. Indeed, this project has led to the development of a glossary of sourdough baking (Roussel et al. (2020)). Participatory research is time consuming as it requires the development of a common vocabulary but also trust between partners. Conducting experiment in a rigorous manner outside the laboratory is also a challenge. It is therefore common to have missing data. Here, in addition to numerous meetings in different locations to design the experiment, the bakers conducted an experiment over several weeks in addition to their professional activity. Their involvement, as well as that of the researchers, made it possible to obtained all the expected data. It should be noted,

398 however, that it is not possible to carry out all possible experiments in participatory research  
399 projects. In particular, in our case, the bakers did not want to carry out sourdough invasion trials  
400 with strains or sourdoughs from elsewhere, fearing that this would alter the microbiota of their  
401 sourdough and their bakery.

## 402 4.2 General features of the sourdoughs

403 The composition of the experimental and collected sourdough microbiota was consistent with  
404 previous studies on sourdough. The mean LAB to yeasts ratio was 65.4, which is within the same  
405 range as that reported by other studies (Zhang et al., 2011; Lhomme et al., 2015; Arici et al., 2017;  
406 Fraberger et al., 2020). As previously detected in French sourdoughs, *F. sanfranscisensis* was the  
407 most frequently encountered bacterial species. *S. cerevisiae*, *K. humilis* and *K. bulderi* were the  
408 most frequently encountered sourdough dominant yeast species (Michel et al., 2016; Urien et al.,  
409 2019; Lhomme et al., 2015). Surprisingly, *Saccharomycetales* accounted for fewer than 5% of the  
410 reads in ten collected sourdough samples, yet a typical yeast density and morphology was observed  
411 in almost all of these samples. This may have reflected biases in the metabarcoding analysis (Loos  
412 and Nijland, 2020). DNA could have been poorly extracted or amplified, thus leading to a low  
413 number of reads. The reads might also not have passed the quality filtering or merging steps in  
414 the bioinformatics analysis, particularly if the ITS region was too long. This is a limitation of the  
415 dada2 software, where reads that are too long to be merged are lost. However, this does not  
416 concern the ITS database, as in this case the ASV would have been found but not assigned to a  
417 species.

### 4.3 Flour-associated species were mainly plant-associated microorganisms

The microbiotas of the experimental and collected flours as well as that of the seeds mainly comprised plant-associated microorganisms. Several filamentous fungi known to be cereal pathogens, and notably *Alternaria* and *Mycosphaerella* species were detected. Similarly, several bacterial genera such as *Pseudomonas* and *Pantoea* were found. Many species in these genera are plant pathogens or plant-associated species (Dutkiewicz et al., 2016; Preston, 2004).

Most of the species that we detected in seeds and flour during this study had been mentioned in previous studies on wheat seed microbiotas (Kuzniar et al., 2020; Rozhkova et al., 2021; Minervini et al., 2018b). They had also been mentioned in studies describing flour microbiota (Landis et al. (2021)), and the results were in accordance with those of De Angelis et al. (2019) who compared the microbiotas of soft and *durum* wheat flour using culture independent methods. Minervini et al. (2018a) analyzed the microbiotas of three different flours, and found the species *F. sanfranciscensis* in every sample (4% of all the strains isolated from the flour). This was higher than what we found, and could have been related to bias affecting the culture independent analyses, where rare species can go undetected, especially when the sequencing depth is low. In the flour samples, most of the reads were discarded as they corresponded to mitochondria or plasts, thus leading to a lower detection of rare bacteria than in sourdough. We did not have this problem for fungi, and the sequencing depth was the same for flour and sourdough, so we can be confident that flours do not contain sourdough fermentative yeasts.

The filamentous fungi plant-associated pathogens detected in flour were also detected in sourdoughs, as they have commonly been detected in other studies (Landis et al. (2021); Reese et al. (2020)). However, on average they accounted for 54% of the reads (sd = 30%) in sourdough and 92% (sd = 9.3%) in flour, suggesting that filamentous fungi die in the acidic environment of sourdough and/or are poor competitors with yeasts in this environment. To our knowledge, they

443 have never been detected alive in sourdough, even though they are able to grow on the media  
444 classically used to enumerate yeasts (Me and Melvydas, 2007). The presence of their DNA in  
445 sourdough suggested that this was partly protected in this environment, possibly thanks to their  
446 cell wall structure. However the high proportion of these fungi in sourdough may also be related  
447 to bias affecting DNA extraction and amplification.

448 Unlike filamentous fungi and several bacteria of the core microbiota, the common plant bacteria  
449 *Pseudomonas* was not detected in sourdoughs, suggesting they did not survive in the sourdough  
450 ecosystem and that their DNA was degraded. This is highly probable as *Pseudomonas* species  
451 generally do not survive at a low pH.

#### 452 **4.4 Microbial succession during the establishment of new sourdough**

453 During the first week of sourdough initiation, LAB of the genera *Pediococcus* and *Weisella* were  
454 detected. They were then replaced with species of the family Lactobacillaceae, such as *Fructi-*  
455 *lactobacillus sanfranciscensis* or *Companilactobacillus paralimentarius*, in accordance with other  
456 studies on sourdough initiation (Bessmeltseva et al., 2014). None of the Ascomycota yeast species  
457 found in newly initiated sourdough was detected in flour and grains used, suggesting that sour-  
458 dough yeasts don't come from the flour but rather from the house microbiota. In the experiments,  
459 the bakers used spoon to mix flour and water. Therefore, the air and utensils are more likely the  
460 vector of microorganisms than their hands.

#### 461 **4.5 Microorganism present in flour did not develop in mature sour-** 462 **dough**

463 The analysis of 46 flours collected all over France confirmed that flour does not contain fermenting  
464 yeasts found in mature sourdoughs.

465 In addition, our results showed that mature sourdoughs did not contain the same LAB as

466 those provided by the flour. *F. sanfranciscensis*, which is the most frequently encountered LAB  
467 species in sourdough, was almost never found in flour. The most abundant *F. sanfranciscensis*  
468 ASV in sourdoughs, which is shared across all the French sourdoughs studied, was never detected  
469 in flour samples. We cannot exclude the fact that this ASV was missed in the flour because  
470 the detection threshold of our metabarcoding analysis was not sufficient. However, conversely,  
471 rare *F. sanfranciscensis* ASVs were detected in five flour samples, but were not found in the  
472 sourdough backslopped with these flours except in one case. It therefore seems that the few  
473 *F. sanfranciscensis* strains present in the flour do not establish in the sourdough. This result is  
474 contradicted by a previous study (Minervini et al., 2018a) which showed that flour and sourdough  
475 share the same strains of *F. sanfranciscensis* in three bakeries.

#### 476 **4.6 Yeast distribution among sourdoughs**

477 Previous work showed that *K. humilis* and *K. bulderi* were commonly found in sourdoughs from  
478 french bakers. Their distribution was correlated with the type of bread-making practices. Sour-  
479 dough made by farmer bakers tended to carry *K. bulderi* while sourdough made according to  
480 artisanal practices often contained *K. humilis* (Michel et al., 2019). One of the main difference  
481 between these types of bakers is that farmer bakers exchange seeds, share mills or supply each  
482 other with flour, while artisanal bakers usually buy their flour from millers who produce and store  
483 flour at a larger scale. So one hypothesis explaining this yeast distribution might be that different  
484 sources of flour supply may lead to different pathways for microorganism dispersal and explain the  
485 structuring of yeast species diversity as a function of bread-making practices. In the present study,  
486 none of the yeast species usually found in sourdough was detected in flour indicating that it is  
487 unlikely that sourdough yeasts come from the flour. The preferential occurrence of *K. bulderi* in  
488 sourdoughs made by farmer-bakers or *K. humilis* in artisanal sourdoughs cannot therefore be easily  
489 explained by differences in flour and wheat grains supply chains. This finding is in agreement with  
490 previous studies which showed that the species composition of sourdough yeasts depended more

491 on the bakery house than on the cereal flour species used (Minervini et al., 2015; Comasio et al.,  
492 2020). The yeast distribution has to be explained by another factor, such as for example different  
493 bread-making practice that may lead to selection of different species. Sourdough yeast may also  
494 disperses as a result of exchanges of sourdough between bakers and/or through meetings in the  
495 bakery where bakers could exchange utensils, handshakes and contribute to carry the sourdough  
496 microorganism. Additional experiments should be carried out to test for these hypothesis.

497 In conclusion, the evaluation of the bacterial and fungal composition of wheat grain, flour and  
498 sourdough showed that microbial species present in the flour are mostly not found in the sourdough.  
499 They do not develop in mature sourdough and therefore are not being part of the microbiota active  
500 during bread-making process. As a corollary, dispersal from the house microbiota and selection  
501 by the sourdough ecological niche appear as the main drivers of the mature sourdough microbial  
502 composition.



## Acknowledgments

We would like to thank H el ene Chaudy, Lili Moyses, Xavier Dell'Armi, Michel Perrin, Florent Mercier, Gilles Simonneaux, Laurent Marteau for their active participation to the project, as well as 47 other bakers and farmer-bakers that have shared their sourdough, flour and knowledges. More generally, we would like to thank the farmers association "Triptol eme", Biocivam11" and "R eseau Semences Paysannes" that contributed to the dynamics of the participatory project. We are also grateful to Sylvain Santoni and Audrey Weber for the Illumina sequencing and their valuable advices as well as Matthieu Barret for his advice for the study of grains microbiota. We also would like to thank Julien Lebrat and Simon Rousselot for their technical assistance. Finally we would like to thank Dominique Desclaux and Kristel Moinet that supervised the project (Gluten: mythe ou r ealit e ?) which allowed the collection and analysis of a part of the flour and mature sourdough samples. This work was supported by a grant from the French National Research Agency (ANR-13-ALID-0005) and a grant from the Fondation de France (Gluten: mythe ou r ealit e ?).

## References

- Arici, M., Ozulku, G., Yildirim, R. M., Sagdic, O., and Durak, M. Z. (2017). Biodiversity and technological properties of yeasts from Turkish sourdough. *Food Sci Biotechnol*, 27(2):499–508.
- Arora, K., Ameer, H., Polo, A., Di Cagno, R., Rizzello, C. G., and Gobbetti, M. (2021). Thirty years of knowledge on sourdough fermentation: A systematic review. *Trends in Food Science & Technology*, 108:71–83.
- Bigey, F., Segond, D., Friedrich, A., Guezenc, S., Bourgeois, A., Huyghe, L., Agier, N., Nidelet, T., and Sicard, D. (2020). Evidence for two main domestication trajectories in *Saccharomyces cerevisiae* linked to distinct bread-making processes. *bioRxiv*, page 2020.05.28.120584.
- 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. *Nature Methods*, 13(7):581–583.
- Calvert, M. D., Madden, A. A., Nichols, L. M., Haddad, N. M., Lahne, J., Dunn, R. R., and McKenney, E. A. (2021). A review of sourdough starters: Ecology, practices, and sensory quality with applications for baking and recommendations for future research. *PeerJ*, 9:e11389.

- 532 Carbonetto, B., Ramsayer, J., Nidelet, T., Legrand, J., and Sicard, D. (2018). Bakery yeasts, a  
533 new model for studies in ecology and evolution. *Yeast*, 35(11):591–603.
- 534 Comasio, A., Verce, M., Van Kerrebroeck, S., and De Vuyst, L. (2020). Diverse Microbial  
535 Composition of Sourdoughs From Different Origins. *Front Microbiol*, 11.
- 536 [dataset] Michel, E. (2016). Microbial community dispersal in sourdough - data and scripts.  
537 <https://www.ebi.ac.uk/ena/browser/view/PRJEB52442?show=reads>.
- 538 [dataset] von Gastrow, L. (2021). Microbial community dispersal in sourdough - sequences.  
539 <https://www.ebi.ac.uk/ena/browser/view/PRJEB52423?show=reads>.
- 540 De Angelis, M., Minervini, F., Siragusa, S., Rizzello, C. G., and Gobbetti, M. (2019). Wholemeal  
541 wheat flours drive the microbiome and functional features of wheat sourdoughs. *International*  
542 *Journal of Food Microbiology*, 302:35–46.
- 543 De Vuyst, L., Harth, H., Van Kerrebroeck, S., and Leroy, F. (2016). Yeast diversity of sour-  
544 doughs and associated metabolic properties and functionalities. *International Journal of Food*  
545 *Microbiology*, 239:26–34.
- 546 Dutkiewicz, J., Mackiewicz, B., Lemieszek, M. K., Golec, M., and Milanowski, J. (2016). *Pantoea*  
547 *agglomerans* : A mysterious bacterium of evil and good. Part III. Deleterious effects: Infections  
548 of humans, animals and plants. *Ann Agric Environ Med.*, 23(2):197–205.
- 549 Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., Maman, S.,  
550 Hernandez-Raquet, G., Combes, S., and Pascal, G. (2018). FROGS: Find, Rapidly, OTUs  
551 with Galaxy Solution. *Bioinformatics*, 34(8):1287–1294.
- 552 Fraberger, V., Unger, C., Kummer, C., and Domig, K. J. (2020). Insights into microbial diversity  
553 of traditional Austrian sourdough. *LWT*, 127:109358.
- 554 Gänzle, M. and Ripari, V. (2016). Composition and function of sourdough microbiota: From  
555 ecological theory to bread quality. *International Journal of Food Microbiology*, 239:19–25.
- 556 Gänzle, M. G. and Zheng, J. (2019). Lifestyles of sourdough lactobacilli – Do they matter for  
557 microbial ecology and bread quality? *International Journal of Food Microbiology*, 302:15–23.
- 558 Gobbetti, M., Minervini, F., Pontonio, E., Di Cagno, R., and De Angelis, M. (2016). Drivers for  
559 the establishment and composition of the sourdough lactic acid bacteria biota. *International*  
560 *Journal of Food Microbiology*, 239:3–18.
- 561 Kuzniar, A., Włodarczyk, K., Grzadziel, J., Wozniak, M., Furtak, K., Galazka, A., Dziadczyk, E.,  
562 Skorzynska-Polit, E., and Wolinska, A. (2020). New Insight into the Composition of Wheat  
563 Seed Microbiota. *International Journal of Molecular Sciences*, 21(13):4634.

- 564 Landis, E. A., Oliverio, A. M., McKenney, E. A., Nichols, L. M., Kfoury, N., Biango-Daniels, M.,  
565 Shell, L. K., Madden, A. A., Shapiro, L., Sakunala, S., Drake, K., Robbat, A., Booker, M.,  
566 Dunn, R. R., Fierer, N., and Wolfe, B. E. (2021). The diversity and function of sourdough  
567 starter microbiomes. *eLife*, 10:e61644.
- 568 Lau, S. W., Chong, A. Q., Chin, N. L., Talib, R. A., and Basha, R. K. (2021). Sourdough  
569 Microbiome Comparison and Benefits. *Microorganisms*, 9(7):1355.
- 570 Lhomme, E., Lattanzi, A., Dousset, X., Minervini, F., De Angelis, M., Lacaze, G., Onno, B., and  
571 Gobbetti, M. (2015). Lactic acid bacterium and yeast microbiotas of sixteen French traditional  
572 sourdoughs. *International Journal of Food Microbiology*, 215:161–170.
- 573 Liu, Z., Lozupone, C., Hamady, M., Bushman, F. D., and Knight, R. (2007). Short pyrosequencing  
574 reads suffice for accurate microbial community analysis. *Nucleic Acids Res*, 35(18):e120–e120.
- 575 Loos, L. M. and Nijland, R. (2020). Biases in bulk: DNA metabarcoding of marine communities  
576 and the methodology involved. *Mol Ecol*, page mec.15592.
- 577 McMurdie, P. J. and Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive  
578 Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, 8(4):e61217.
- 579 Me, V. and Melvydas, V. (2007). Primary analysis of new measures against fungal diseases of  
580 woody plants. *Biologija*, 18(1):4.
- 581 Meroth, C. B., Walter, J., Hertel, C., Brandt, M. J., and Hammes, W. P. (2003). Monitoring the  
582 Bacterial Population Dynamics in Sourdough Fermentation Processes by Using PCR-Denaturing  
583 Gradient Gel Electrophoresis. *Appl. Environ. Microbiol.*, 69(1):475–482.
- 584 Michel, E., Masson, E., Bubbendorf, S., Lopicque, L., Legrand, J., Guézenec, S., Nidelet, T.,  
585 Marlin, T., Rué, O., Onno, B., Sicard, D., and Bakers, T. P. (2019). Artisanal and farmers  
586 bread making practices differently shape fungal species diversity in French sourdoughs. *bioRxiv*,  
587 page 679472.
- 588 Michel, E., Monfort, C., Deffrasnes, M., Guezenec, S., Lhomme, E., Barret, M., Sicard, D.,  
589 Dousset, X., and Onno, B. (2016). Characterization of relative abundance of lactic acid bacteria  
590 species in French organic sourdough by cultural, qPCR and MiSeq high-throughput sequencing  
591 methods. *Int. J. Food Microbiol.*, 239:35–43.
- 592 Mikkelsen, G. M. (2005). Niche-Based vs. Neutral Models of Ecological Communities. *Biol Philos*,  
593 20(2-3):557–566.
- 594 Minervini, F., Dinardo, F. R., Celano, G., De Angelis, M., and Gobbetti, M. (2018a). Lactic Acid  
595 Bacterium Population Dynamics in Artisan Sourdoughs Over One Year of Daily Propagations  
596 Is Mainly Driven by Flour Microbiota and Nutrients. *Front. Microbiol.*, 9.
- 597 Minervini, F., Lattanzi, A., De Angelis, M., Celano, G., and Gobbetti, M. (2015). House mi-  
598 crobiotas as sources of lactic acid bacteria and yeasts in traditional Italian sourdoughs. *Food*  
599 *Microbiology*, 52:66–76.

- 600 Minervini, F., Lattanzi, A., De Angelis, M., Di Cagno, R., and Gobbetti, M. (2012). Influence of  
601 Artisan Bakery- or Laboratory-Propagated Sourdoughs on the Diversity of Lactic Acid Bacterium  
602 and Yeast Microbiotas. *Applied and Environmental Microbiology*, 78(15):5328–5340.
- 603 Minervini, F., Lattanzi, A., Dinardo, F. R., De Angelis, M., and Gobbetti, M. (2018b). Wheat  
604 endophytic lactobacilli drive the microbial and biochemical features of sourdoughs. *Food Mi-*  
605 *crobiology*, 70:162–171.
- 606 Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel,  
607 D., Kennedy, P., Picard, K., Glöckner, F. O., Tedersoo, L., Saar, I., Kõljalg, U., and Abarenkov,  
608 K. (2019). The UNITE database for molecular identification of fungi: Handling dark taxa and  
609 parallel taxonomic classifications. *Nucleic Acids Res*, 47(D1):D259–D264.
- 610 Preston, G. M. (2004). Plant perceptions of plant growth-promoting *Pseudomonas*. *Philos Trans*  
611 *R Soc Lond B Biol Sci*, 359(1446):907–918.
- 612 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner,  
613 F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and  
614 web-based tools. *Nucleic Acids Res*, 41(Database issue):D590–D596.
- 615 Ravi, R. K., Walton, K., and Khosroheidari, M. (2018). MiSeq: A Next Generation Sequencing  
616 Platform for Genomic Analysis. In DiStefano, J. K., editor, *Disease Gene Identification: Methods*  
617 *and Protocols*, Methods in Molecular Biology, pages 223–232. Springer, New York, NY.
- 618 Reese, A. T., Madden, A. A., Joossens, M., Lacaze, G., and Dunn, R. R. (2020). Influences of  
619 Ingredients and Bakers on the Bacteria and Fungi in Sourdough Starters and Bread. *mSphere*,  
620 5(1).
- 621 Roussel, P., Onno, B., Michel, E., and Sicard, D. (2020). *La panification au levain naturel*.  
622 éditions Quae.
- 623 Rozhkova, T. O., Burdulanyuk, A. O., Bakumenko, O. M., Yemets, O. M., Vlasenko, V. A.,  
624 Tatarynova, V. I., Demenko, V. M., Osmachko, O. M., Polozhenets, V. M., Nemerytska, L. V.,  
625 Zhuravska, I. A., Matsyura, A. V., and Stankevych, S. V. (2021). Influence of seed treatment on  
626 microbiota and development of winter wheat seedlings. *Ukrainian Journal of Ecology*, page 7.
- 627 Scheirlinck, I., der Meulen, R. V., Vuyst, L. D., Vandamme, P., and Huys, G. (2009). Molecular  
628 source tracking of predominant lactic acid bacteria in traditional Belgian sourdoughs and their  
629 production environments. *Journal of Applied Microbiology*, 106(4):1081–1092.
- 630 Tamang, J. and Kailasapathy, K. (2010). *Fermented Foods and Beverages of the World*. CRCNET  
631 Books. CRC Press.
- 632 Team, R. C. (2019). R : A Language and Environment for Statistical Computing. R Foundation  
633 for Statistical Computing.

- 634 Thompson, P. L., Guzman, L. M., De Meester, L., Horváth, Z., Ptacnik, R., Vanschoenwinkel,  
635 B., Viana, D. S., and Chase, J. M. (2019). A process-based metacommunity framework linking  
636 local and regional scale community ecology. Preprint, Ecology.
- 637 Urien, C., Legrand, J., Montalent, P., Casaregola, S., and Sicard, D. (2019). Fungal Species  
638 Diversity in French Bread Sourdoughs Made of Organic Wheat Flour. *Front Microbiol*, 10.
- 639 Van Kerrebroeck, S., Maes, D., and De Vuyst, L. (2017). Sourdoughs as a function of their  
640 species diversity and process conditions, a meta-analysis. *Trends in Food Science & Technology*,  
641 68:152–159.
- 642 Vellend, M. (2010). Conceptual Synthesis in Community Ecology. *The Quarterly Review of*  
643 *Biology*, 85(2):183–206.
- 644 von Gastrow, L. (2021). Microbial community dispersal in sourdough - data and scripts.  
645 <https://data.inrae.fr/dataset.xhtml?persistentId=doi:10.15454/DF0BRL>.
- 646 White, T. J., Bruns, T., Lee, S., and Taylor, J. (1990). 38 - AMPLIFICATION AND DIRECT  
647 SEQUENCING OF FUNGAL RIBOSOMAL RNA GENES FOR PHYLOGENETICS. In Innis,  
648 M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J., editors, *PCR Protocols*, pages 315–322.  
649 Academic Press, San Diego.
- 650 Zhang, J., Liu, W., Sun, Z., Bao, Q., Wang, F., Yu, J., Chen, W., and Zhang, H. (2011). Diversity  
651 of lactic acid bacteria and yeasts in traditional sourdoughs collected from western region in Inner  
652 Mongolia of China. *Food Control*, 22(5):767–774.

## 653 **Data accessibility**

654 Raw data and scripts are available at :

655 <https://data.inrae.fr/dataset.xhtml?persistentId=doi:10.15454/DF0BRL>

656 The sequencing data generated in this study were deposited in the European Nucleotide Archive  
657 (ENA) under accession numbers PRJEB52423 for the 46 collected sourdoughs [dataset] von Gas-  
658 trow (2021) and PRJEB52442 for the experimental sourdough initiation [dataset] Michel (2016).

## 659 **Conflict of interest**

660 The authors declare no conflict of interest

## 661 **Author contributions**

662 L. von Gastrow, E. Michel, J. Legrand, P. Rouseel, V. Chable, S. Rousselot, I. Goldringer, X.  
663 Dousset, E. Serpolay-Bessoni, B. Taupier-Letage, C. Vindras-Fouillet, B. Onno, F. Valence and  
664 D. Sicard designed research, L. von Gastrow, E. Michel, R. Amelot, M. Deffrasnes, B. Onno, D.  
665 Sicard, D. Segond, I. Goldringer and S. Guézennec performed research, L. von Gastrow, E. Michel,  
666 O. Rué, J. Legrand, B. Onno and D. Sicard analyzed data and L. von Gastrow, E. Michel, F.  
667 Valence and D. Sicard wrote the paper.

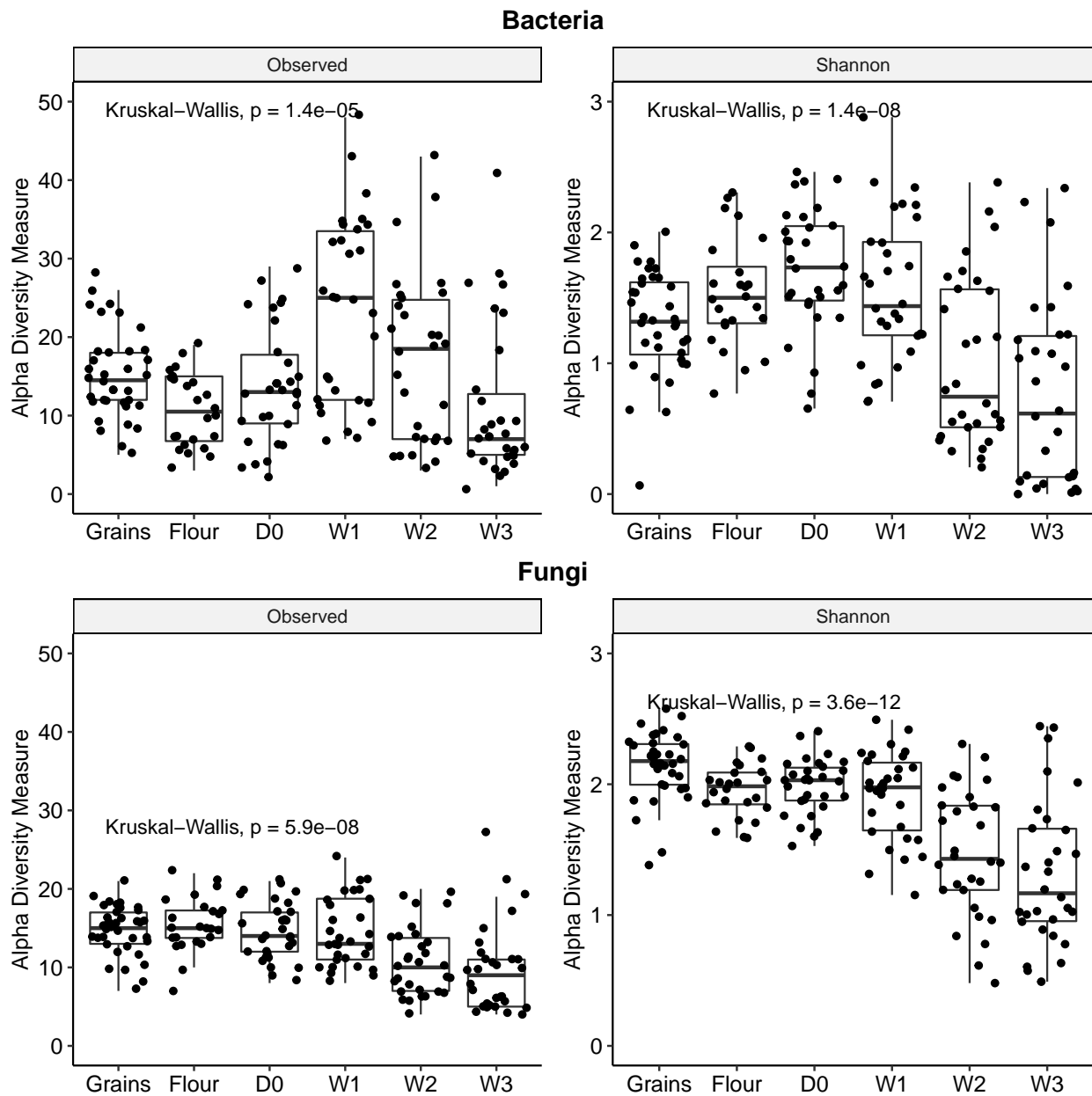


Figure 1: Alpha diversity in sourdough and flour samples, estimated from 16S V3-V4 and ITS1 Illumina MiSeq reads assigned to species. Species richness (on the left) and evenness (on the right) are plotted for grains, flours, and sourdough at their initiation (D0) and after one, two and three weeks of back-slopping (W1, W2 and W3). Data concern bacteria at the top and fungi at the bottom.

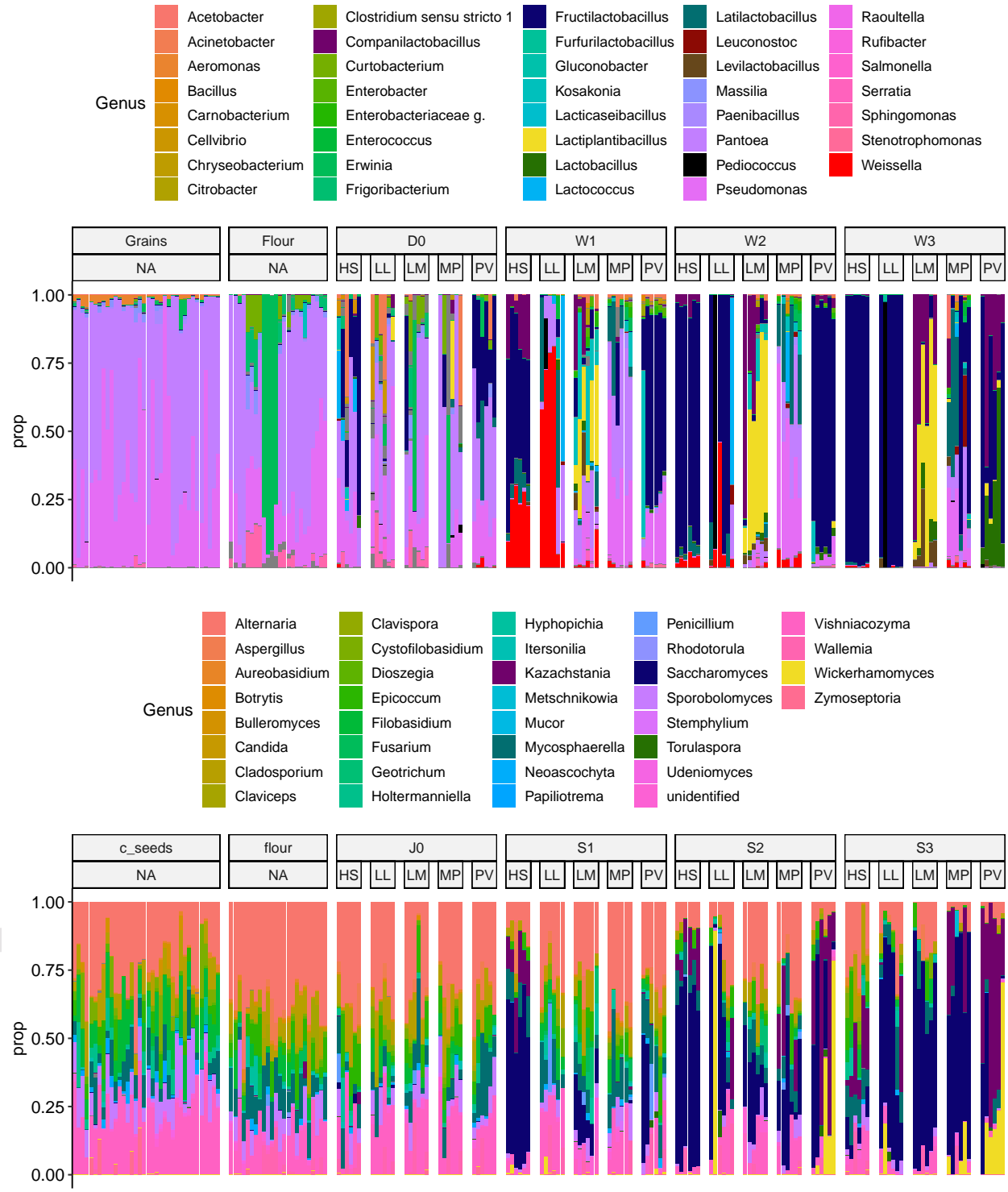


Figure 2: Microbiota along the sourdough making food chain. Seeds, flours and newly initiated sourdoughs samples are plotted on the x axis, and their proportion in the sample (based on the unrarefied data) on the y axis. Colors represent genus. The horizontal white bars delimit different ASV. The four bakers (HS, LM, MP, PV) and the lab produced six sourdoughs out of the six batches of flour produced. The sourdoughs were sequenced at their initiation (J0), and once a week (S1, S2, S3). Each flour was sequenced in duplicate at the beginning and at the end of the experiment, resulting in sequencing of 24 samples of flour.



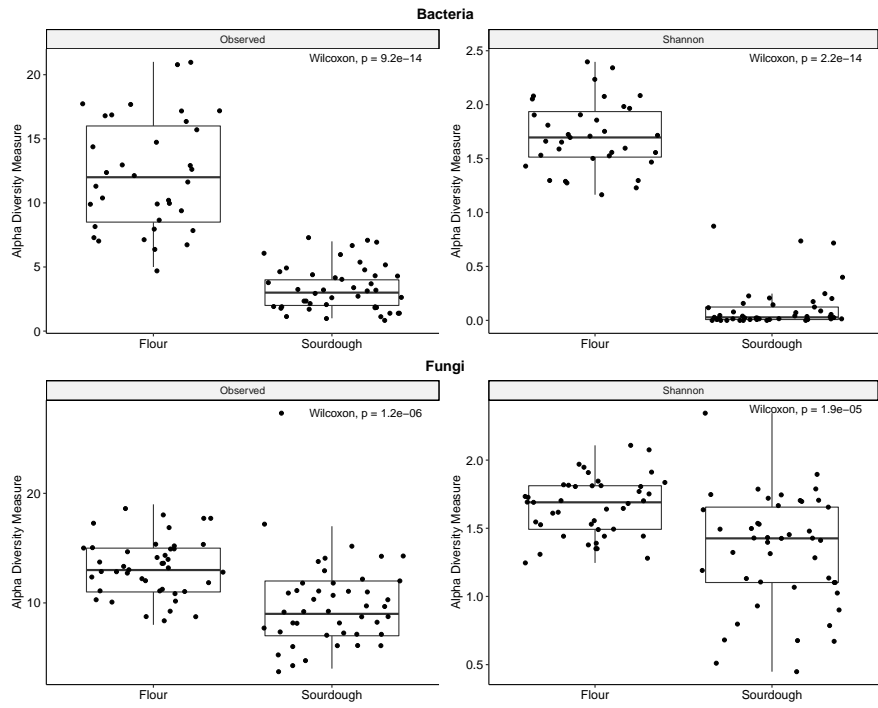


Figure 3: Alpha diversity in sourdough and flour samples, estimated from 16S V3-V4 and ITS1 Illumina MiSeq reads assigned to species. Species richness (on the left) and evenness (on the right) are plotted.

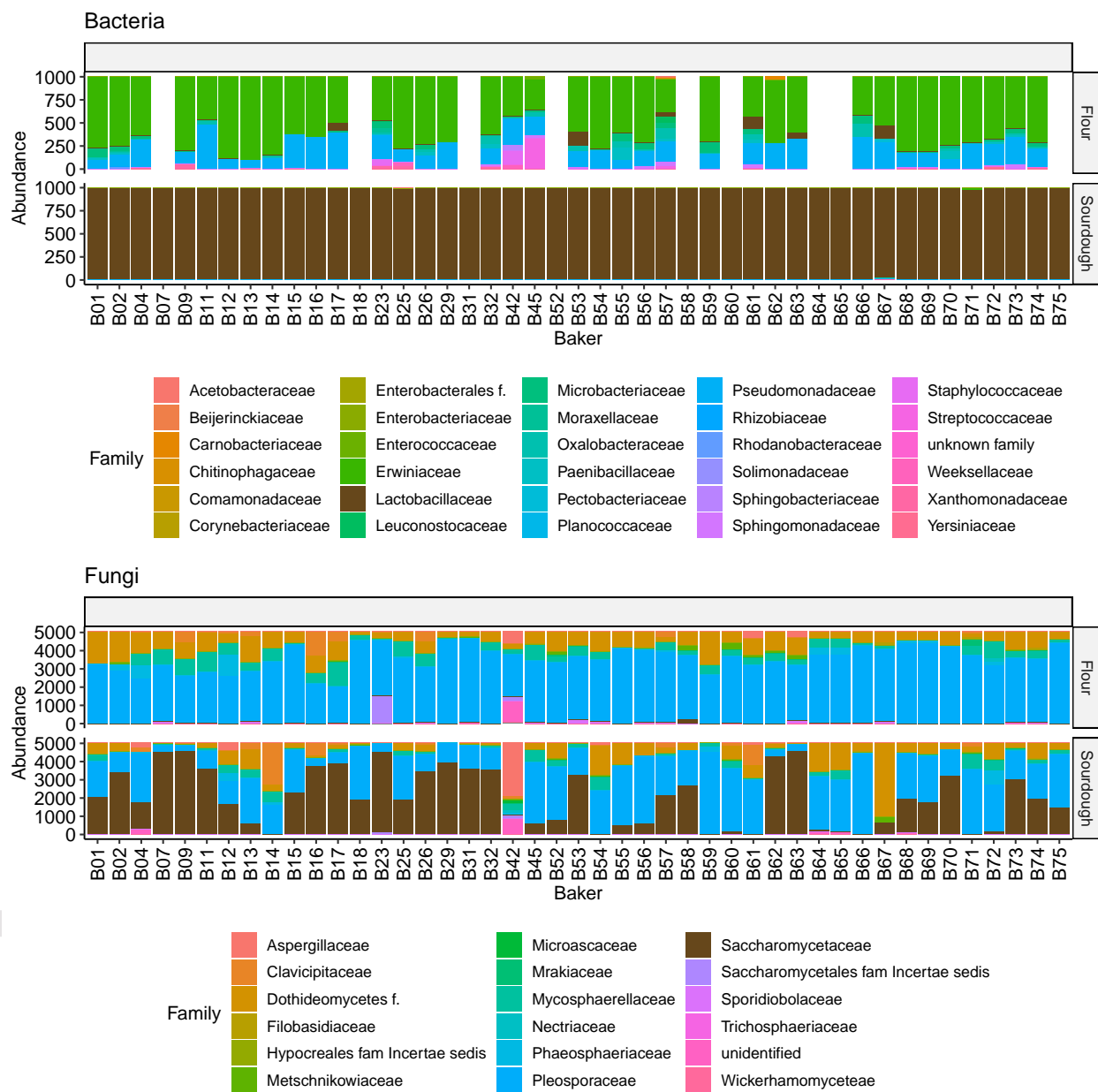


Figure 4: Microbial composition of the collected flours and sourdoughs. Colors represent families, their abundance based on the rarefied data is plotted. Horizontal white bars delimit the different ASVs. Data are missing when samples contained less than 1000 reads due to rarefaction, resulting in no data for eight flours.

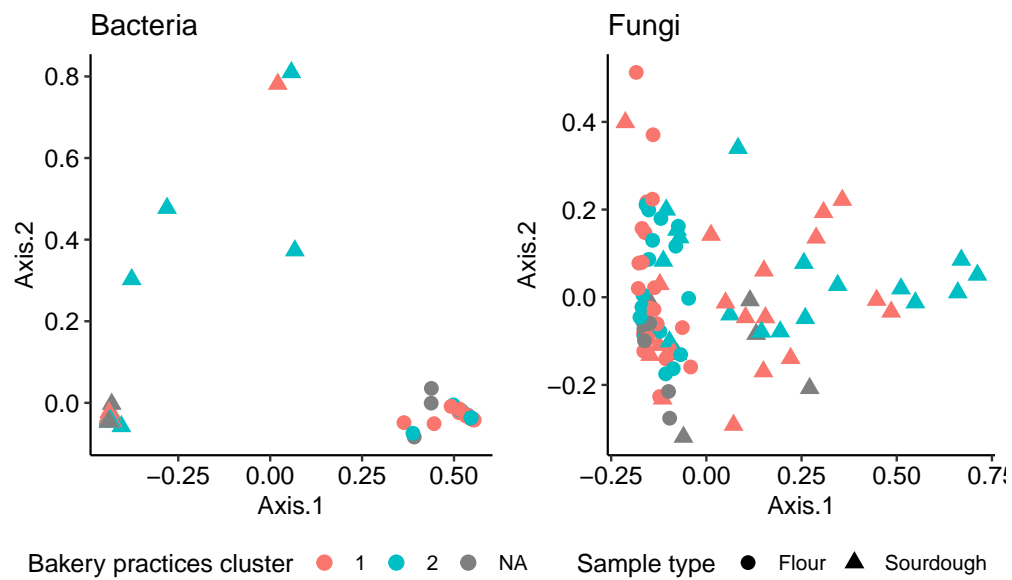


Figure 5: PCoA based on Bray-Curtis dissimilarity for bacteria (left) and fungi (right). Bray-Curtis dissimilarity was computed on the basis of the abundance of the different species. Each point represents a sample. Colors indicate the bakery practices cluster, with farmer-baker practices in red and artisan-baker practices in blue. Sample types are represented by different shapes, flours being shown as circles and sourdoughs as triangles.

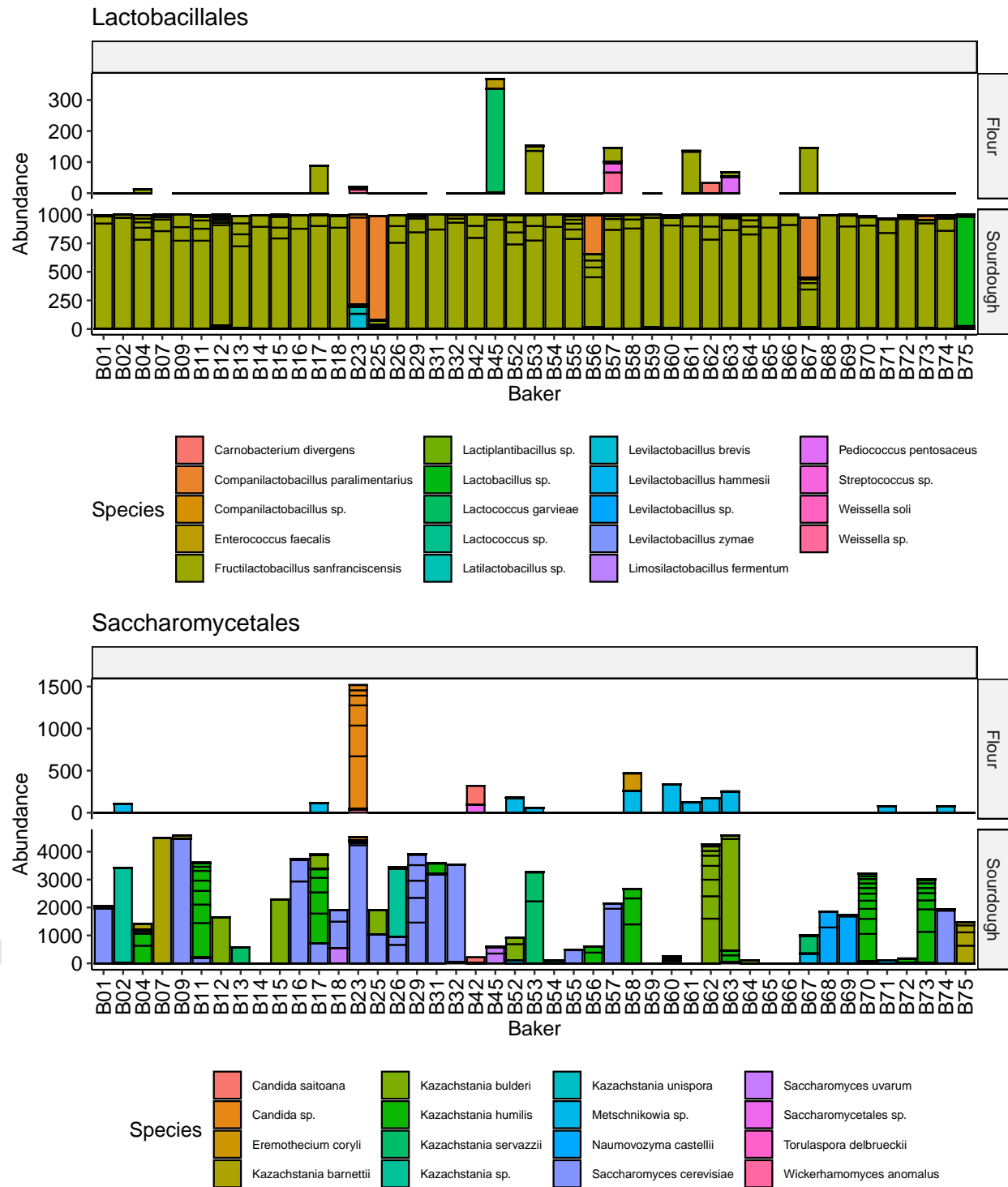


Figure 6: Microbial composition of the collected flours and sourdoughs. Samples are plotted on the x axis, and their abundance based on the rarefied data on the y axis. Colors represent species. Horizontal white bars delimit the different ASVs. Only *Lactobacillales* and *Saccharomycetales* are shown. The axes have different scales for abundance in flour and sourdough.

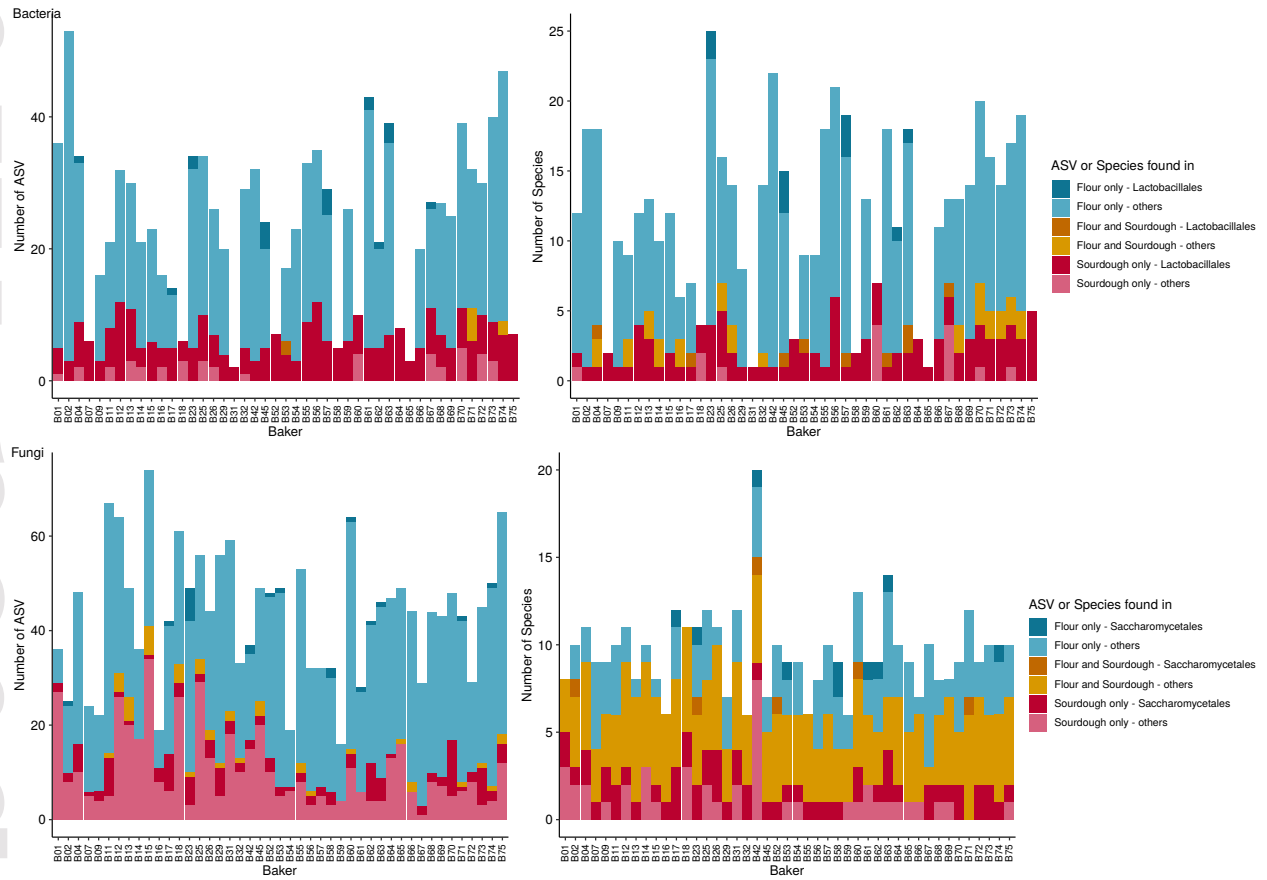


Figure 7: Number of shared species (on the right) and ASV (on the left) between sourdoughs and the flour used to make them. Results for bacteria are shown at the top and for fungi at the bottom.