Title: Artisanal and farmers bread making practices differently shape fungal species diversity in French sourdoughs

Running title: Bread making practices as a driver of yeast species diversity

Authors

Elisa Michel^{*,†}, Estelle Masson[‡], Sandrine Bubbendorf[‡], Léocadie Lapicque[‡], Judith Legrand[§], Stéphane Guézenec[†], Thibault Nidelet[†], Thérèse Marlin[†], Olivier Rué[¶], Bernard Onno^{*}, Delphine Sicard[†] and the participating bakers[#]

Authors affiliation

*Oniris, Laboratoire MicrobioTech, UMR GEPEA 6144, Rue de la Géraudière CS 82225, 44322 Nantes Cedex 3, France

†SPO, Inra, Univ. Montpellier, Montpellier SupAgro, 2 place Pierre Viala, 34090 Montpellier, France

‡Laboratoire de Psychologie : Cognition, Comportement, Communication – EA 1285, Université de Bretagne Occidentale, 20 rue Duquesne, CS 93837, F-29238 Brest 03, France

§Génétique Quantitative et Evolution-Le Moulon, INRA, Université Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

¶MaIAGE, INRA, Université Paris-Saclay, Jouy-en-Josas, France

#Bakers and farmers

Corresponding author Delphine Sicard, SPO, Inra, Univ. Montpellier, Montpellier SupAgro, Montpellier, France +33 4 99 61 24 60 delphine.sicard@inra.fr

No competing interests

Financial supports: French National Research Agency (ANR-13-ALID-0005) and INRA

Abstract

Preserving microbial diversity in food systems is one of the many challenges to be met to achieve food security and quality. However, there is still a lack of knowledge about the factors that may influence food microbial diversity, notably in fermented foods. Although industrialization led to the selection and spread of specific fermenting microbial strains, there are still ongoing artisanal processes that may allow the conservation of a wider diversity of microbial species. We examined whether the diversity of artisanal practices could lead to an increased level in fungal species diversity for bread making. We used an interdisciplinary participatory research approach including bakers, psychosociologists and microbiologists to analyse French bread making practices and describe fungal communities in naturally fermented sourdough. Bread making practices were clustered in a farmer practices' group and an artisanal practices' group. Surprisingly, the well-known bakery yeast, *Saccharomyces cerevisiae*, was dominant (i.e. with a relative abundance over 50%) in only 24% of sourdoughs and other yeast species of the closely related *Kazachstania* genus were frequent. Bread making practices were found to drive the distribution of these species. The differences in fungal communities were associated with variation in sourdough acidity, maltose concentration and hydration. Overall, our results showed that preserving bread making practices diversity allows the preservation of a higher taxonomic and functional diversity in microbial communities.

Introduction

Humans started to ferment food before the emergence of the Neolithic using naturally fermenting microbial communities. In the 19th century, the industrialization and the increase of knowledge in microbiology resulted in changes in fermented food practices with the use of targeted species and selected strains. This selection led to a reduction in species and genetic diversity for fermented food processing and limited *in situ* conservation of microbial communities in industrialized systems. Domestication of the yeast *Saccharomyces cerevisiae* for the production of beer, wine, cheese, leavened bread or that of the fungus *Penicillium roquefortii* for blue cheese production are well studied cases [1–5]. Recently, in response to global changes, there is a renewed interest in artisanal practices

that make use of naturally fermenting microbial communities. However, the effect of artisanal practices on the distribution of microbial species diversity remains poorly documented. Thus, understanding whether and how artisanal practices may promote microbial diversity is a first step toward the development of food systems with improved sustainability.

Among fermented foods, bread is still a symbol deeply engrained in the history, religious rites and medicine of several cultures. The origin of bread likely originated 14 000 years ago, suggesting bread was made long before plant domestication [6]. Since the Neolithic, bread history is intimately combined with the evolution of cereals, bread making associated tools and the advent of Mediterranean civilizations [7]. Leavened bread was traditionally made with flour, water and a fermenting agent, which was either a fermenting beverage or a fermenting dough, termed sourdough [1, 8]. This sourdough was generally initiated from a mixture of flour and water, naturally colonized by lactic acid bacteria (LAB) and yeasts [1, 9]. Sourdough was then either maintained from one bread making process to the other or initiated again and again, depending on the craftsman [10, 11]. In the 19th century, yeast starters made of *S. cerevisiae*, often called « baker's yeast », were proposed as an alternative to sourdough and these are nowadays commonly used by local bakeries and industries. However, there is currently a renewed interest in the use of naturally fermented sourdoughs for bread leavening.

In France, sourdough breads are made both by bakers and farmer-bakers who also grow and mill their own wheat. The number of farmers-bakers has increased in the 2000s with two motives: to grow wheat varieties meeting their needs and to assert their independence from industry [12]. Although farmer-bakers are less numerous than bakers, they participate in the renewed interest in local wheat varieties and artisanal know-how, which may contribute to the conservation of both socio-cultural diversity and microbial diversity.

Here, we used a participatory research approach involving psycho-sociologists, microbiologists, bio-mathematicians, bakers and farmers-bakers to study whether and how bakers and farmers-bakers contribute to the preservation of socio-cultural and fungal species diversity in sourdough microbial

community. This research emerged from a co-construction by bakers, farmer-bakers and academic researchers.

Materials and methods

A questionnaire survey, face-to-face interviews and focus-groups to collect bread making practices

All materials were collected among bakers and farmer-bakers through a questionnaire survey, interviews and focus groups. The collected variables were related to *i*) the ingredients origin : wheat varieties types (ancient populations also called landraces / modern varieties), whether they produced flour from their own wheat, whether they had their own mill or use an external mill, water origin, *ii*) the sourdough recipe: its age, its hydration state, the origin of the chief sourdough (sample of dough or sourdough), the number of back-sloppings before bread making and per week, the temperature of water used for back-sloppings, *iii*) their bread making practices: the number of bread makings per week, the percentage of sourdough, flour and salt in bread dough, the kneading methods, the total duration of fermentation and the addition of baker's yeast in dough.

Sourdoughs samples, enumeration and strain isolation

Sourdoughs were collected in 39 French bakeries before kneading and referred to as final sourdoughs (Table S1). On the day of collection, they were sent to the lab where yeast and bacteria were enumerated and isolated as in [13, 14], and sourdoughs stored at -20°C in sterile vials for non- culture based analysis. Ethics and rights associated with sourdough collection and strains isolation have been respected.

Sourdough acidity and metabolic analyses

For each sourdough, three independent 1-g replicates were analysed. pH and Total Titrable Acidity were measured as described in (22). Organic acids, alcohol and sugars concentrations (expressed as g/kg of sourdough) were analysed by liquid chromatography using an HPLC HP 1100 LC system (Agilent technologies, Santa clara, CA, USA) equipped with a refractive index detector (RID Agilent G1382A)

and a UV detector (Agilent G1314A). Two different columns were used, a Rezex ROA-organic acids column and a Rezex RPM-monosaccharide column (SDVB – Pb+2 8%, 300x7.8mm, Phenomenex, Torrance, CA, USA). The details of the procedure are described in supplementary information (Method S1).

Yeast species identification

The Internal transcribed spacer 1 (ITS1) ribosomal DNA of each 1216 yeast isolates was amplified by PCR from chromosomal DNA, either by using primers ITS1F and ITS2 [15, 16], or primers NSA3 and 58A2R [16, 17]. For isolates unidentified with the ITS1 region alone, DNA was extracted according to MasterPure Yeast DNA purification kit (Epicentre, Epibio). PCR reactions targeting partial genes, the D1D2 region of the large subunit of rRNA (LSU), a part of the RNA polymerase II large subunit encoding gene (*RPB1*), a part of the RNA polymerase II encoding gene (*RPB2*), a part of the actin encoding gene *Act1* and Transcriptional Elongation Factor TEF were performed. To discriminate three specific isolates, PCR on genes *GHD1*, *FSY1*, *URA3*, *DRC1*, *MET2* were performed [18–21] (Table S4). All PCR products were sent to be sequenced with Sanger sequencing (Eurofins, Germany). Species were identified using NCBI [22], YeastIP [23] and a personal database, which was constructed after *ITS1*, *RPB2*, LSU sequencing of all 33-yeast species reportedly found in sourdoughs in the literature [1].

Sourdough DNA extraction, MiSeq sequencing, bioinformatics

The ITS1 region was targeted with the PCR primers ITS1-F (5'- CTTGGTCATTTAGAGGAAGTAA - 3') and ITS2 (5'-GCTGCGTTCTTCATCGATGC-3').

The sequencing run was performed with MiSeq Reagent Kit v3. 2015 [13, 24]. Sequences were analysed through FROGS “Find Rapidly OTU with Galaxy Solution” [28] and home-made pipelines. Overlapped reads were merged with Flash [29] with a minimum overlap of 10 nucleotides, a maximum overlap of 300 nucleotides and a maximum mismatch density of 0.1. Adapters were removed with Cutadapt [30] and data were cleaned with Sickle [25] Reads were clustered with Swarm [31] and chimeras deleted with VSEARCH [32]. Sequences were then filtered on minimum abundance of 0.05% of all sequences.

From the OTU's abundance table and for each OTU, the taxonomic affiliation using UNITE Version 7.1, Release 2016-11-20 [33], YeastIP [23] and our own databases [13] [24] was obtained by blasting OTUs representative sequences against each database.

Data analyses

All statistics and plots (ggplot2 [34], leaflet package [35], with minor esthetical adjustment with Inkscape) have been done with R.

To analyse bread making practices, a multiple correspondence analysis (MCA) and hierarchical clustering on principal components based on the first two axes of the MCA were performed using the FactoMineR R package [36].

To analyse fungal community, weighted Unifrac distances between sourdough communities were computed from a rooted phylogenetic tree based on the OTUs sequences using the R-packages Phyloseq and GUniFrac [26, 27, 38]. Phylogenetic sequences were aligned with Clustalo [37] and phylogenetic tree were built with the parsimony algorithm, with 100 replicates bootstraps, pairwise ktuple-distances with Seaview [37]. Different roots were tested (*Sporidiobolales* sp., *Bullera globospora*, *Trichosporon asahii*, *Udeniomyces pyricola*). Tree architecture did not change with the root. It did not fit the expected phylogeny and, notably, some *Ascomycota* were misclassified among the *Basidiomycota*. However, the dominant sourdough species belonging to the *Saccharomycetaceae* family were clustered according to expected clades or subclades, except that *Kazachstania servazzi* and *Kazachstania unispora* were grouped in a clade closer to *Saccharomyces* species than to other *Kazachstania* species. Using the Unifrac distances matrix, we performed a Principal Coordinate Analysis (PCoA) and clustered sourdough communities using the first two axes of the PCoA. To check the sensitivity of our analysis to this misclassification, we performed the same analyses without the sourdoughs that had one misclassified species representing more than 10% of their reads, *i.e.* sourdoughs B20, B41, B42, and B44 and found the same clustering [27].

For each sourdough, the species richness, Chao1, Shannon and Simpson indexes were computed, and the Shannon and Simpson indices values were converted to the effective number of species per sourdough. This number was estimated from the Shannon diversity index as $e^{\frac{1}{H}}$ and from the Simpson diversity index as $\frac{1}{1-Simpson\ index}$ [39, 40]. For probability estimates, the exact 95% confidence intervals were computed using a binomial distribution.

To study the link between microbial diversity and bakery practices, a univariate Permutational Multivariate Analysis (PERMANOVA) on the Unifrac distance matrix was performed for each bakery practice variable. We performed univariate analysis on the 30 bakers who had less than 8 missing values among the 29 bread making practices variables and adjusted the p-value using FDR correction to account for multiple testing. In addition, independence exact Fisher tests between the variable providing fungal community PCoA groups and each of the bread making practices variables were performed. Multiple testing was accounted for using the False Discovery Rate method [41].

The link between the baker practices group, the fungal community group or the yeast dominant species and the variation of each quantitative variable (microbial density, pH, TTA, metabolite concentration) was tested with the following mixed effect model: $Y_{ijk} = \mu + \alpha_i + B_j + \varepsilon_{ijk}$ with $\varepsilon_{ijk} \sim N(0, \sigma^2)$, where α_i is the effect of the fungal community group i modelled as a fixed effect and B_j is the effect of sourdough j modelled as a random effect and k represents the measurement replicates. For sourdough hydration rate, the variable was arcsin transformed but sourdough effect was not included in the model because no repetition was obtained from any sourdough. The model parameters were estimated using the lmerTest R package [42]. To test the fixed effects, we used likelihood ratio tests. Multiple comparisons of means were performed using Tukey tests with the multcomp package [43]. p-values were all adjusted for multiple testing with the FDR method. The geographical structuration was tested with a Mantel test on the Unifrac distances matrix and the geographical distances matrix computed with the package geosphere [44] and ade4 [45].

Data and scripts are shared on Zenodo. <http://doi.org/10.5281/zenodo.2600170>

181

182 **Results and discussion**

183 **Bread making processes and the diversity of bread making practices**

184 We worked with thirty-nine French bakers producing natural sourdough bread and distributed all over
 185 France (Table S1). Their bread making practices were analysed through personal interviews (n=12),
 186 focus groups (n=3), observation during bread-making workshops (n=2), and an online/phone survey
 187 (n=36). The general process of sourdough bread making is presented in Figure 1. We analysed 29
 188 variables around this general bread making process, describing variations of the practices at all steps of
 189 the bread making process, from wheat grains to baked bread (Figure S1). Three bakers who did not
 190 answer the survey on bread making practices, and four others who did not provide enough information
 191 about their practices were excluded from the analysis. According to a hierarchical clustering on principal
 192 components (HCPC), the 32 other bakers clustered into two groups corresponding to two main types of
 193 bread making practices (Figure 2). The first group, hereafter termed “farmer” practices group, includes
 194 6 bakers and 11 farmers-bakers using the following practices: low bread production (<500 kg per week,
 195 81% of the bakers of the “farmer” group), use of ancient wheat populations (56%), manual kneading
 196 (63%), working at ambient temperature (88%), long fermentation periods (more than 4 hours for 88%),
 197 and no use of commercial baker’s yeast (88%). In addition, they tend to make their chief sourdough
 198 from dough after kneading (75%). The second group, hereafter called “artisanal” practices group
 199 consists of 12 bakers and 4 farmer-bakers having more intensive practices, characterized by a large bread
 200 production (>500 kg per week, 81%), mechanical kneading (100%), use of modern wheat varieties
 201 (63%), working at ambient temperature (56%), using commercial yeast starters in addition to sourdough
 202 for bread making or using commercial yeast starters for pastries and buns making (81%). In this second
 203 group, bakers tend to make their chief sourdough from a final sourdough.

Fungal community composition

Sourdough is a mix of flour and water naturally fermented by bacteria and yeasts. Sourdough yeast density ranged from $8.1 \cdot 10^4$ to $5.8 \cdot 10^8$ CFU per gram of sourdough, with a mean value of $2.9 \cdot 10^7$ CFU per gram, as commonly found in sourdoughs from all over the world [1, 17, 46, 47]. We isolated 20 to 40 yeast strains from each sourdough and identified species using ITS sequence as well as other barcodes when needed (see M&M section). A total of 1216 strains were characterized from 38 sourdoughs. In addition, we developed an ITS1 meta-barcoding MiSeq sequencing method on sourdough (see sup M&M). After filtering 5,360,620 raw ITS1 sequences for quality, abundance (0.005%) and chimera, 3,542,801 sequences were further analyzed. Overall, the sequences clustered in 113 OTUs. The number of reads per sourdough ranged from 8421 to 194 557. Therefore, we carried out our analysis on the rarefied matrix. Among all OTUs, 64 were identified as non-yeast, including 10 assigned to the order *Triticodae* (especially to the species *T. aestivum*), 50 assigned to plant pathogen fungi, such as *Alternaria*, *Aspergillus* or *Fusarium*, *Gibberella*, while 4 OTUs remained unidentified. Among the 40 yeast or yeast-like OTUs, 96% of total reads were assigned to the phylum *Ascomycota*, 87.5% to the order *Saccharomycetales* and 85.7% to the family *Saccharomycetaceae*. Only 4% of the total reads were assigned to the phylum *Basidiomycota*. Overall, three OTUs assigned to the species *Kazachstania humilis*, *Kazachstania bulderi* and *Saccharomyces cerevisiae* represented 20.3%, 15.5% and 24.1% respectively of the total number of reads and 28.1%, 23.7% and 18.2% respectively of the number of reads identified as yeast species (Figure 3).

Both non culture-based and culture-based methods allowed the identification of the same dominant species (defined as a species with an over 50% frequency) for all sourdoughs but five (B09, B20, B22, B25, B41) (Figure 3). In two cases, the discrepancy was explained by the detection of *Cladosporium* sp. at high frequency with metabarcoding while this species could not be isolated in the laboratory (Figure 3). In two other cases, it was explained by a high number of *S. cerevisiae* isolated in the laboratory compared to what was observed using metabarcoding sequencing. Finally, in the last case, the identification of *Pichia kudriavzevii* required additional sequencing as it shares an identical ITS with *Candida xylopsoci*. Because metabarcoding allows a deeper characterization of the fungal species

diversity with few discrepancy cases, the distribution of fungal species diversity will be further described using metabarcoding data only. Previous analysis of the same sourdoughs revealed that *L. sanfranciscensis* was the dominant bacterial species in all analysed sourdoughs but two, where the dominant species was either *L. curvatus* or *L. heilongjiangensis* [13, 14]. Therefore, we decided to study the distribution pattern of microbial species in the fungal community only.

Distribution of fungal species diversity over sourdoughs

All sourdoughs but two had a dominant yeast species with a relative abundance over 50% and many species with a lower relative abundance (Figure 3). Within sourdoughs, fungal species richness ranged from 10 to 33, with a 23.5 median (Table S2). The effective number of species per sourdough calculated from the Shannon diversity index ranged from 1 to 6.8 (Table S2), with 70% of sourdoughs having an index below 2 (Table S2). Between-sourdoughs species diversity was analysed using weighted Unifrac distances, computed from a phylogenetic tree built from the distances between OTUs using *Sporidiobolales* species as root (Figure S2). Unifrac distances computed with four differently rooted trees were highly positively correlated (Figure S2). Unifrac distances between sourdoughs ranged from 0.0005 and 0.71, with a median of 0.49 and a mean of 0.52. There was no significant correlation between the Unifrac distances and geographical distances between sourdoughs (Mantel test, $P=0.35$) (Figure 4).

We then analysed specifically the distribution of yeast species diversity as yeast species, together with lactic acid bacteria, are the main functional player in a sourdough ecosystem and for bread quality. Over the 40 yeast species detected in the 38 sourdoughs, 12 had a relative abundance over 50% in at least one sourdough, four had a relative abundance between 20% and 50% and 24 had a relative abundance below 10%. All dominant species (relative abundance over 50%) were fermentative yeast species, except in one sourdough that had a *Cladosporium* species. We found all the sourdough yeast genera (*Saccharomyces*, *Candida*, *Kazachstania*, *Pichia*, *Torulaspora* and *Hyphopichia*) commonly reported in the literature except the *Wickerhamomyces* genus that we did not detect in our samples [47–49].

The baker's yeast, *Saccharomyces cerevisiae* is not the most widespread yeast species in sourdoughs

The well-known baker's yeast, *Saccharomyces cerevisiae*, was found in 53% of all sourdoughs (95% confidence intervals=36% - 69%) but was dominant (relative abundance over 50%) in only 24% (95% confidence intervals=11% - 40%) (Figure 3). In two cases, *S. cerevisiae* co-occurred with another yeast species at similar relative abundance. In the first case, *S. cerevisiae* was present at a relative abundance of 40% with *Candida sake* at a 41% relative abundance. In the second case, it was found at a relative abundance of 47% with *Pichia kudriavzevii* at a relative abundance of 52%. In all the other cases, *S. cerevisiae* had a relative abundance below 21% and was found with other dominant yeast species, such as *Kazachstania australis*, *Kazachstania humilis*, *Saccharomyces uvarum* or *Torulaspora delbrueckii*. This suggests that *S. cerevisiae* did not displace other species and can indeed be out-competed by other species in sourdoughs. Therefore, despite the recurrent use of *S. cerevisiae* as starter, its massive use in some bakeries and homes, and its occurrence in a wide range of habitats such as soil, trees, and humans [2, 3], this species does not appear to have overwhelmingly colonized French traditional sourdoughs (Figure 3).

The *Kazachstania* genus is highly represented in sourdoughs

Kazachstania was the most represented yeast genus over all sourdoughs, when considering both the number of reads over all sourdoughs and the number of detected species. Indeed, this genus represented 57% of the total number of reads while *Saccharomyces* represented 26% of the total number of reads. In addition, eight species of the *Kazachstania* genus were found in sourdough, a much higher figure than for the *Saccharomyces* genus, represented by only two species (*S. uvarum* and *S. cerevisiae*) (Figure 3). *Kazachstania* species dominated in 54% (95% confidence intervals=36%-69%) of sourdoughs while *Saccharomyces* species dominated in 27% only (95% confidence intervals=13%-43%). *K. humilis*, followed by *K. bulderi* were the most commonly dominant *Kazachstania* species, and found in respectively 21% (95% confidence intervals=10%-37%) and 15% of sourdoughs (95% confidence intervals=6%-31%) (Figure 3). *K. humilis* is common in sourdoughs and has been found in many

countries, viz. China, Ethiopia, Finland, Germany, Morocco, USA, Italy, Belgium and France [49–52]. *K. bulderi* has been reported for the first time in anaerobic maize silage in the Netherlands and in fermented liquid feed for piglets [53, 54] and more recently in French wheat sourdoughs [14, 17]. A yet undescribed *Kazachstania* species was also identified in five sourdoughs (4.5%-29%) and found dominant in three (1.7%-22%). Strains of this species were closely related to a strain previously isolated from boza, a Bulgarian fermented drink, as estimated with ITS and LSU (D1D2) barcodes (Source: NCBI, GenBank: KC118125.1 and KX369579.1). In addition, *Kazachstania saulgeensis*, a recently described species [55, 56], was dominant in one sourdough (0.07%-14%). Finally, several *Kazachstania* species were detected for the first time as dominant in sourdoughs, whereas they had been previously found in other environments, like soil (*K. australis*), sauerkraut (*K. barnettii*), fermented milk (*K. unispora*), or feces (*K. unispora*) [55, 57, 58]. None of the previous studies on sourdough have evidenced as many *Kazachstania* species [1].

The diversity of sourdough fungal communities was associated with differences in bread making practices

We tested whether sourdough fungal community composition could be explained by bread making practices. To do so, we performed univariate PERMANOVA analysis on the 30 bakers with less than 8 missing values for the 29 bread making practices variables (Table S3). The univariate analysis revealed that the weighted Unifrac distance between sourdoughs varied significantly ($P < 0.05$) with the use of commercial yeast in bakery. It also varied significantly with sourdough age, chief sourdough origin (dough, sourdough or both), the quantity of bread produced per week, the milling method (cylinder, millstone, Astrie, Tyrol), the type of wheat variety (ancient, modern or a mix thereof) and the fermentation duration. However, after FDR correction for taking into account multiple testing, none of these variables significantly explained Unifrac distances.

In order to understand further the relationship between sourdough fungal community composition and bread making practices, we clustered sourdoughs according to their fungal community composition, on the basis of the PCoA of their weighted Unifrac distances. Then, we tested the link between the

fungal community group and the bread making practice group (farmers/artisanal practices group) as well as the link between the fungal community group and each of the different bread making practices (Figure 5). Sourdoughs were clustered into three fungal community groups. Group 1 clustered all sourdoughs (but two) having *Kazachstania* species as dominant species (*K. humilis*, *K. barnettii*, *K. bulderi*, *K. saulgeensis*, *K. sp.*). Group 2 contained sourdoughs with *Saccharomyces sp.*, *K. servazzi* or *K. unispora* as dominant species. Group 3 sourdoughs harbored *S. cerevisiae* together with other species such as *Pichia kudriavzevii*, *Candida sake*, or a *Dipodascaceae sp.* Group 1 sourdoughs were mostly made by bakers having farmer's bread making practices while group 2 and 3 sourdoughs were mostly made by bakers using artisanal practices (exact Fisher test, $P=0.035$). The fungal community groups were significantly related to two specific bread making practice variables: the quantity (in kg) of bread made per week (Exact Fisher test, $P=0.001$) and the use of commercial yeast (Exact Fisher test, $P=0.05$). All sourdoughs in group 2 but one were found in bakeries making between 500 kg and 1000 kg of bread per week, while groups 1 and 3 sourdoughs originated from bakeries producing very different amounts of bread (ranging from amounts below 250 kg to over 1000 kg). In addition, group 1 sourdoughs were more frequently found in bakeries that do not use commercial yeast while group 2 and 3 were more frequently found in bakeries using the commercial yeast *S. cerevisiae* (Exact Fisher test, $P=0.01$). Interestingly, group 1 sourdoughs harbored *S. cerevisiae* either at a relative abundance below 1% or not at all, while all groups 2 and 3 sourdoughs had *S. cerevisiae* at a relative abundance over 20%, except in three cases where it was either absent or at a relative abundance below 6%.

To test more specifically the link between bread making practices and the distribution of *Kazachstania* species, we analyzed more in depth group 1 sourdoughs. Within this group, 8 sourdoughs had *K. humilis* as dominant species, 6 had *K. bulderi*, 3 had the undescribed *Kazachstania* species and the remainder had still other *Kazachstania* species. All sourdoughs made with artisanal practices carried *K. humilis* as dominant species or, in one case, the yet undescribed *Kazachstania sp.* By contrast, sourdoughs made with farmers' practices had as dominant species *Kazachstania bulderi*, *K. australis*, *K. barnettii*, *K. saulgeensis* or the yet undescribed *Kazachstania* species (exact Fisher test, $P=0.004$).

Other bread making practices than the ones studied here could also explain the distribution of yeast species diversity. Interviews with the 5 bakers working with sourdough hosting *Kazachstania bulderi* and the underscribed *Kazachstania* species suggested the role of dispersion of these species in French sourdoughs. Indeed, these bakers have been connected over the years either through seed exchanges, sourdoughs mixing or gifts, bread making training in common or working in one another's bakery. Although sourdough bacteria have been shown to originate from the bakery house microbiota and flour [59, 60], the origins of yeast species found in sourdough are yet unknown.

Fungal community composition was partly related to sourdough acidity, maltose concentration and hydration

The composition of fungal community may affect sourdough metabolic content (sugars, acids, alcohols) via fungal strains metabolite consumption and production. Inversely, the presence and concentration of different compounds (sugars, acids, alcohols) may also be one of the driver of fungal communities' composition as those parameters may affect differently the different fungal strains fitness. For example, lactic acid bacteria (LAB) are the main producers of acidity in sourdough, but yeasts also produce acetic acid and may also indirectly affect acidity through positive or negative interaction with bacteria.

To investigate the relation between sourdough fungal communities and metabolic compounds, we quantified sourdough hydration, yeast density, bacteria density, sourdough pH, Total Titrable Acidity (TTA), sourdough concentration in seven sugars (maltose, glucose, fructose, raffinose, arabinose, mannose, xylose), four alcohols (glycerol, ethanol, mannitol, meso-erythritol), six acids (lactate, acetate, glutarate, pyruvate, malate, succinate) and calculated the fermentative quotient (lactate over acetate ratio). For each variable, there was a wide range of variation (Table S5). The principal component analysis based on all variables showed no evidence of sourdoughs grouping (Fig S4). As expected in fermentation, yeast density was positively correlated to ethanol ($r=0.74$, $P<0.001$), glycerol ($r=0.67$, $P<0.001$), and acetate ($r=0.6$, $P<0.001$) concentration. However, it was not significantly correlated to

sugar concentrations. This might be explained by sugar competition and cross feeding that may exist between yeasts and bacteria.

We then tested whether bread making practices group (farmers' practices and artisanal practices) was associated with the variation of each quantitative variable separately. There was no significant effect of the bread making practice group except for sourdough hydration that was significantly higher in sourdoughs made using farmers' practices ($F_{1,94}=11.69$, $P<0.001$). Sourdoughs made with farmer's practices had in average 55% water while sourdoughs made with artisanal practices had in average 49% of water.

In addition, we tested whether differences in fungal community were associated with any variation in a quantitative variable (Table S5). Group 3 microbial community sourdoughs (defined by PCoA clustering on Unifrac distance, see below), which contain *S. cerevisiae* in co-dominance with a second yeast species (*Candida sake*, *Pichia kudriavzevii* or a *Dipodascus* species), had a significantly higher mean pH (mean $pH_{\text{group3}}=4.2$ against $pH_{\text{group1}}=3.8$, Tukey Contrasts, $P<0.001$), lower TTA (mean $TTA_{\text{group3}}=7.7$ against $TTA_{\text{group1}}=17.1$, Tukey Contrasts, $P=0.002$), and a higher maltose concentration (mean $Maltose_{\text{group3}}=52.8$ mg/gr of sourdough against $Maltose_{\text{group1}}=24.1$ mg/gr of sourdough, Tukey Contrasts $P=0.002$) than group1, having a *Kazachstania* dominant species. Compared to group 2 having in most cases *S. cerevisiae* as dominant species, it also had higher pH ($pH_{\text{group2}}=3.9$, Tukey Contrasts, $P=0.003$), and higher maltose concentration ($Maltose_{\text{group2}}=23.7$, Tukey contrast, $P=0.003$). These data may reflect a lower fermentative activity for group 3 fungal community having two co-dominant species, and/or a negative interaction effect of group3 fungal community on the activity of lactic acid bacteria (LAB), which are the main producers of sourdough acids. Previous studies on the bacteria content of the same sourdoughs showed that *L. sanfranciscensis* was most generally the dominant species, although *L. heilongjiangensis*, *L. curvatus* or *L. brevis* were also found as dominant species [13, 14]. We found no significant correlation between LAB and yeast densities ($r=-0.15$, $p=0.45$, Fig S4) but the link between fungal and bacteria community might be species and strains dependent. Additional studies on the interactions between fungal and bacterial communities need to be performed to better understand how they may drive sourdough acidity and sugar content.

Finally, we analysed whether the difference in dominant yeast species was associated with some variation of any of the studied quantitative variables. We only considered the 26 sourdoughs having either *S. cerevisiae* (9 sourdoughs), *K. humilis* (8 sourdoughs), *K. bulderi* (6 sourdoughs) or *K. sp* (3 sourdoughs) as dominant species, since the other yeast species were found dominant only once. The differences in dominant species was not significantly associated to variation in sourdough sugar, acids or alcohol concentration. However, *K. bulderi* was found in significantly more hydrated sourdoughs (63% water content in average) compared to the three other dominant species, *K. humilis*, *K. sp.*, and *S. cerevisiae*, found in sourdoughs having respectively 49%, 47%, 53 % water content in average (Tukey Contrasts, $P < 0.001$; $P < 0.001$; $P < 0.001$). *K. bulderi* was found to be dominant only in sourdoughs made using farmers' practices, a bread making practice group that was also found to be associated with more hydrated sourdoughs. Additional experiments should be carried out to test whether this species has indeed a better fitness in more hydrated sourdoughs or whether its presence in more hydrated sourdoughs is related to covariation with other farmer practices.

Conclusion and perspectives

In conclusion, a great diversity of bread making practices and fungal community composition was found in our sample of French sourdoughs. Surprisingly, the well-known baker's yeast *Saccharomyces cerevisiae* was found dominant only in one fourth of the sampled sourdoughs. By contrast, several species of the neighbouring genus *Kazachstania* (including one yet undescribed species) were detected at high frequency, revealing a major role for this mostly unknown genus in the study of fungal domestication and in bread making. The fungal community diversity was partly explained by the diversity in bread making practices. It was also partly associated with sourdough acidity and maltose concentration, suggesting the role of fungi/bacteria interaction in sourdough functional diversity. To our knowledge, this is the first evidence of the influence of artisanal practices on taxonomic and functional diversity in microbial communities. Therefore, our results highlight the necessity of maintaining socio-cultural diversity to maintain microbial diversity in food systems. These findings

could not have been evidenced without the collaboration of bakers and scientists, showing the importance of participatory research projects to gain new insight into biodiversity preservation.

Acknowledgements

This work has emanated from research conducted with the financial support of the French National Research Agency, ANR-13-ALID-0005 and INRA. We thank all the bakers and farmers-bakers for constructing the project with us and for sharing their sourdoughs and knowledge. We thank Yoann Robert, Candice Aulard, Josette Bessière for lab assistance and Matthieu Barret for MiSeq sequencing. We also would like to thank Philippe Roussel for sharing his knowledge about French bread.

References

1. Carbonetto B, Ramsayer J, Nidelet T, Legrand J, Sicard D. Bakery yeasts, a new model for studies in ecology and evolution. *Yeast* 2018.
2. Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 2018; **556**: 339.
3. Legras J-L, Galeote V, Bigey F, Camarasa C, Marsit S, Nidelet T, et al. Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. *Molecular biology and evolution* 2018; **35**: 1712–1727.
4. Ropars J, López-Villavicencio M, Snirc A, Lacoste S, Giraud T. Blue cheese-making has shaped the population genetic structure of the mould *Penicillium roqueforti*. *PloS one* 2017; **12**: e0171387.
5. Independent domestication events in the blue-cheese fungus *Penicillium roqueforti* | bioRxiv. <https://www.biorxiv.org/content/10.1101/451773v1.abstract>. Accessed 21 Mar 2019.
6. Arranz-Otaegui A, Carretero LG, Ramsey MN, Fuller DQ, Richter T. Archaeobotanical evidence reveals the origins of bread 14,400 years ago in northeastern Jordan. *Proceedings of the National Academy of Sciences* 2018; 201801071.
7. Samuel D. Bread in archaeology. *Civilisations Revue internationale d'anthropologie et de sciences humaines* 2002; 27–36.

8. Samuel D. Bread in archaeology. *Civilisations Revue internationale d'anthropologie et de sciences humaines* 2002; 27–36.
9. Gobbetti M, Minervini F, Pontonio E, Di Cagno R, De Angelis M. Drivers for the establishment and composition of the sourdough lactic acid bacteria biota. *International journal of food microbiology* 2016; **239**: 3–18.
10. Roussel P, Chiron H, Paillard G. Les pains français: évolution, qualité, production. 2002. Maë-Erti.
11. Decock P, Cappelle S. Bread technology and sourdough technology. *Trends in Food Science & Technology* 2005; **16**: 113–120.
12. Demeulenaere É, Rivière P, Hyacinthe A, Baltassat R, Baltazar S, Gascuel J-S, et al. La sélection participative à l'épreuve du changement d'échelle. À propos d'une collaboration entre paysans sélectionneurs et généticiens de terrain. *Natures Sciences Sociétés* 2017; **25**: 336–346.
13. Michel E, Monfort C, Deffrasnes M, Guezenec S, Lhomme E, Barret M, et al. Characterization of relative abundance of lactic acid bacteria species in French organic sourdough by cultural, qPCR and MiSeq high-throughput sequencing methods. *International journal of food microbiology* 2016; **239**: 35–43.
14. Lhomme E, Orain S, Courcoux P, Onno B, Dousset X. The predominance of *Lactobacillus sanfranciscensis* in French organic sourdoughs and its impact on related bread characteristics. *International journal of food microbiology* 2015; **213**: 40–48.
15. White TJ, Bruns T, Lee S, Taylor JW, others. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications* 1990; **18**: 315–322.
16. Martin KJ, Rygiewicz PT. Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. *BMC Microbiology* 2005; **5**: 28.
17. Urien C, Legrand J, Montalent P, Casaregola S, Sicard D. Fungal species diversity in French bread sourdoughs made of organic wheat flour. *Frontiers in Microbiology* .

18. Peris D, Sylvester K, Libkind D, Gonçalves P, Sampaio JP, Alexander WG, et al. Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. *Molecular ecology* 2014; **23**: 2031–2045.
19. Nguyen H-V, Gaillardin C. Evolutionary relationships between the former species *Saccharomyces uvarum* and the hybrids *Saccharomyces bayanus* and *Saccharomyces pastorianus*; reinstatement of *Saccharomyces uvarum* (Beijerinck) as a distinct species. *FEMS yeast research* 2005; **5**: 471–483.
20. Libkind D, Hittinger CT, Valério E, Gonçalves C, Dover J, Johnston M, et al. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proceedings of the National Academy of Sciences* 2011; **108**: 14539–14544.
21. Sampaio JP, Gonçalves P. Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. *Applied and environmental microbiology* 2008; **74**: 2144–2152.
22. The NCBI Handbook, 2nd ed. 2013. National Center for Biotechnology Information (US).
23. Weiss S, Samson F, Navarro D, Casaregola S. YeastIP: a database for identification and phylogeny of *Saccharomycotina* yeasts. *FEMS yeast research* 2013; **13**: 117–125.
24. Barret M, Briand M, Bonneau S, Prévieux A, Valière S, Bouchez O, et al. Emergence Shapes the Structure of the Seed Microbiota. *Appl Environ Microbiol* 2015; **81**: 1257–1266.
25. Core UDB. Sickle: A windowed adaptive trimming tool for FASTQ files using quality. UC Davis, California, United States of America URL: <http://bioinformatics.ucdavis.edu/software/> (Accessed: 7 September 2015) 2015.
26. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* 2013; **8**: e61217.
27. Chen J. GUniFrac: generalized UniFrac distances. *R package version* 2012; **1**: 2012.
28. Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, et al. FROGS: find, rapidly, OTUs with galaxy solution. *Bioinformatics* 2017; **34**: 1287–1294.
29. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 2011; **27**: 2957–2963.

30. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 2011; **17**: pp–10.
31. Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* 2014; **2**: e593.
32. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 2016; **4**: e2584.
33. Kõljalg U, Larsson K-H, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, et al. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist* 2005; **166**: 1063–1068.
34. Wickham H. ggplot2: elegant graphics for data analysis. 2016. Springer.
35. Cheng J, Karambelkar B, Xie Y, Wickham H, Russell K, Johnson K. Leaflet: Create Interactive Web Maps with the JavaScript “Leaflet” Library (2017). URL [https://CRAN.R-project.org/package=](https://CRAN.R-project.org/package=leaflet) *leaflet R package version* ; **1**: 423.
36. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *Journal of statistical software* 2008; **25**: 1–18.
37. Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution* 2009; **27**: 221–224.
38. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology* 2005; **71**: 8228–8235.
39. Chao Anne, Chiu Chun-Huo, Jost Lou. Phylogenetic diversity measures based on Hill numbers. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2010; **365**: 3599–3609.
40. Jost L. Partitioning diversity into independent alpha and beta components. *Ecology* 2007; **88**: 2427–2439.
41. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society Series B (Methodological)* 1995; 289–300.

42. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 2017; **82**.
43. Hothorn T, Bretz F, Westfall P, Heiberger RM, Schuetzenmeister A, Scheibe S, et al. Package ‘multcomp’. See <https://cran.r-project.org/web/packages/multcomp/index.html> 2017.
44. Hijmans RJ, Williams E, Vennes C, Hijmans MRJ. Package ‘geosphere’. 2017.
45. Dray S, Dufour A-B. The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software* 2007; **22**: 1–20.
46. Clément H, Prost C, Chiron H, Ducasse MB, Della Valle G, Courcoux P, et al. The effect of organic wheat flour by-products on sourdough performances assessed by a multi-criteria approach. *Food Research International* 2018; **106**: 974–981.
47. De Vuyst L, Van Kerrebroeck S, Leroy F. Microbial Ecology and Process Technology of Sourdough Fermentation. *Advances in Applied Microbiology* 2017.
48. Siepmann FB, Ripari V, Waszczynskyj N, Spier MR. Overview of Sourdough Technology: from Production to Marketing. *Food and Bioprocess Technology* 2017; 1–29.
49. Van Kerrebroeck S, Maes D, De Vuyst L. Sourdoughs as a function of their species diversity and process conditions, a meta-analysis. *Trends in Food Science & Technology* 2017; **68**: 152–159.
50. Lhomme E, Urien C, Legrand J, Dousset X, Onno B, Sicard D. Sourdough microbial community dynamics: An analysis during French organic bread-making processes. *Food Microbiology* 2016; **53, Part A**: 41–50.
51. Wang X, Du H, Zhang Y, Xu Y. Environmental Microbiota Drives Microbial Succession and Metabolic Profiles during Chinese Liquor Fermentation. *Applied and environmental microbiology* 2018; **84**: e02369–17.
52. Valmorri S, Tofalo R, Settanni L, Corsetti A, Suzzi G. Yeast microbiota associated with spontaneous sourdough fermentations in the production of traditional wheat sourdough breads of the Abruzzo region (Italy). *Antonie Van Leeuwenhoek* 2010; **97**: 119–129.
53. Middelhoven WJ, Kurtzman CP, Vaughan-Martini A. *Saccharomyces bulderi* sp. nov., a yeast that ferments gluconolactone. *Antonie van Leeuwenhoek* 2000; **77**: 223–228.

54. Gori K, Bjørklund MK, Canibe N, Pedersen AØ, Jespersen L. Occurrence and identification of yeast species in fermented liquid feed for piglets. *Microbial ecology* 2011; **61**: 146–153.
55. Jacques N, Sarilar V, Urien C, Lopes MR, Morais CG, Uetanabaro APT, et al. Three novel ascomycetous yeast species of the Kazachstania clade, *Kazachstania saulgeensis* sp. nov., *Kazachstania serrabonitensis* sp. nov. and *Kazachstania australis* sp. nov. Reassignment of *Candida humilis* to *Kazachstania humilis* fa comb. nov. and *Candida pseudohumilis* to *Kazachstania pseudohumilis* fa comb. nov. *International journal of systematic and evolutionary microbiology* 2016; **66**: 5192–5200.
56. Sarilar V, Sterck L, Matsumoto S, Jacques N, Neuvéglise C, Tinsley CR, et al. Genome sequence of the type strain CLIB 1764T (= CBS 14374T) of the yeast species *Kazachstania saulgeensis* isolated from French organic sourdough. *Genomics Data* 2017; **13**: 41–43.
57. Kurtzman CP. Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and other members of the *Saccharomycetaceae*, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygorhizula*. *FEMS yeast research* 2003; **4**: 233–245.
58. Boynton PJ, Greig D. The ecology and evolution of non-domesticated *Saccharomyces* species. *Yeast* 2014; **31**: 449–462.
59. Minervini F, Lattanzi A, De Angelis M, Celano G, Gobbetti M. House microbiotas as sources of lactic acid bacteria and yeasts in traditional Italian sourdoughs. *Food Microbiology* 2015; **52**: 66–76.
60. Minervini F, Lattanzi A, Dinardo FR, De Angelis M, Gobbetti M. Wheat endophytic lactobacilli drive the microbial and biochemical features of sourdoughs. *Food microbiology* 2018; **70**: 162–171.

Figure Legends

Figure 1. The sourdough bread making process. Sourdough is a mix of flour and water naturally fermented by bacteria and yeasts. It is initiated by mixing flour, water and occasionally other ingredients. It is then fed by regularly adding flour and water, a process termed back-slopping. Once considered mature by bakers based on their acidity, flavour and bubbling activity, the sourdough is called "chief", or "mother" sourdough, and can then be used for bread making. The bread making process starts from this "chief sourdough", or from a piece of dough or sourdough sampled from the preceding bread making process, or initiated from a mix of flour and water naturally colonized by yeasts and LAB following several back-sloppings. The chief sourdough is refreshed to constitute the final sourdough, which is used for bread making. This final sourdough is mixed with flour, water, and other ingredients (salt, seeds, yeasts starters, etc.) during kneading to constitute the dough. After kneading, primary fermentation occurs during the first rising. The dough is then divided and shaped. The pieces of dough are then left to rise during a second fermentation and finally oven-baked.

Figure 2. Multiple Correspondence Analysis (MCA) based on 28 categorical variables describing bread making practices.

A) Representation of bakers. Each point represents a bakery. The left area brings together baker with "artisanal" practices and the right area bakers with "farmer" practices. Centroid of each group is designed by a triangle. The dot's colors indicate the PCoA cluster of the sourdough fungal community. Black dots for group 1, white for group 2, grey for group 3. The fungal community of the sourdough of baker 14 (in red) was not studied. B) Representation of the 20 first categories that mostly contributed the MCA axis. The category is written next to the triangle. C) Percentage of different categories in each bread making practices group. Only variables that mostly explained differences between bread making practice groups are shown: use of commercial yeast, kneading method, chief origin, kg of bread production per week, number of bread making per week, percent of water in dough, number of back-sloppings before making bread, water origin, sourdough age and flour percentage in dough

Figure 3. Yeast species diversity was analyzed for 38 out of the 39 sourdoughs with both cultural and metabarcoding methods. Sourdough B14 was excluded from microbial analyses, because of difficulties in isolating strains from the sample, suggesting that this sourdough's microflora was no longer active. Left: species were identified by traditional microbial isolation and identification using ITS sequencing. Right: species were identified using ITS1 metabarcoding.

Figure 4. Distribution of yeast species diversity across French sourdough. Each bar represents the yeast species diversity of one sourdough and is placed on the map where the baker is located.

Figure 5. Representation of fungal community weighted Unifrac distances between sourdoughs. The first line shows the clustering of sourdough according to their Unifrac distance on a tree (left) or on a PCoA (right). Sourdough fungal community can be clustered in three groups according to their weighted Unifrac distances. The lines below show the distribution of bread making practices or the distribution of the dominant or most frequent species among fungal community. Each line corresponds to a variable. On the left: a tree constructed from the Unifrac distance matrix colored by groups of sourdough defined by the clustering analysis. In the center: sourdoughs represented on the first 2 axes of the PCoA colored according to the group of sourdough. On the right: the distribution of modalities of the variable for each group of sourdough (number of each group on the x-axis).

Chief sourdough

Back-sloppings
+ Water, Flour

Final sourdough

+ Water, Flour, Salt

Kneading

First Rising
(Fermentation)

Shaping

Final Rising (Fermentation)

Baking

A)

B)

C)

Unifrac tree, clustered in 3 groups

PCoA, clustering in 3 groups

Practices clustering

Dominant species

Use of commercial yeast

Quantity of bread per week