Title: Artisanal and farmer bread making practices differently shape fungal species community composition in French sourdoughs

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

Authors

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

Authors affiliation

[†] SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier, France

° Oniris, Laboratoire MicrobioTech, UMR GEPEA 6144, Rue de la Géraudière CS 82225, 44322 Nantes Cedex 3, 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, Université Paris-Sud, INRAE, CNRS, AgroParisTech, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

[¶]Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France

[¶]Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

*Judith Legrand and Delphine Sicard should be considered both last authors.

Corresponding author

Delphine Sicard, SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier, France +33 624720501

delphine.sicard@inrae.fr

Key words: domestication, yeast, *Kazachstania*, *Saccharomyces cerevisiae*, fermentation

No competing interests

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

Abstract

Preserving microbial diversity in food systems is one of the many challenges to be met to achieve food security and quality. 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 species diversity and genetic diversity. 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, psycho-sociologists and microbiologists to analyze French bread making practices and describe fungal communities in naturally fermented sourdough of 27 bakers and 12 farmer bakers. Bread making practices were classified in two groups: the farmer-like practice group and the artisanal-like practice group. The well-known bakery yeast, *Saccharomyces cerevisiae,* was dominant (i.e. with a relative abundance over 50%) in only 24% of sourdoughs while other yeast species, belonging to the *Kazachstania* genus, were dominant in 54% of sourdoughs. Bread making practices were found to drive the distribution of fungal species across sourdoughs. The most striking bread making practice effect was the occurrence of *Kazachstania humilis* in sourdoughs made with artisanal-like practices and the occurrence of *Kazachstania bulderi* in sourdoughs made with farmer-like practices. Phenotypic divergences between sourdough and non-sourdough strains were found for *K. humilis* but not for *K. bulderi*. Overall, our results showed that preserving bread making practice diversity allows the preservation of a higher species and phenotypic diversity in microbial communities.

57 # Introduction

58 Humans started to ferment food before the Neolithic using naturally fermenting microbial
59 communities. In the 19th century, the industrialization and the increase of knowledge in microbiology
60 resulted in changes in fermented food practices with the use of starters. This selection led to a reduction
61 in species diversity and genetic diversity for fermented food processing and limited *in situ* conservation
62 of microbial communities in industrialized systems [1–3]. Domestication of the yeast *Saccharomyces*
63 *cerevisiae* for the production of beer, wine, cheese, leavened bread, that of the fungi *Penicillium*
64 *roqueforti* or *Penicillium camemberti* for cheese production or that of the fungus *Aspergillus oryzae* for
65 rice or soybean fermented products are well studied cases [1–10]. The recent renewed interest in
66 artisanal practices that make use of naturally fermenting microbial communities could promote the
67 conservation of microbial diversity. However, the effect of artisanal practices on the distribution of
68 microbial species across sourdoughs remains poorly documented.

69 Among fermented foods, bread is still a symbol deeply engrained in the history, religious rites
70 and medicine of several cultures. Bread likely originated 14 000 years ago, suggesting that bread was
71 made long before plant domestication [11]. Since the Neolithic, bread history is intimately associated
72 with the domestication of cereals, bread making associated tools and the advent of Mediterranean
73 civilizations [12]. Investigation of the morphology of plant remains which were incorporated in
74 Neolithic bread identified wheat, barley, millet, linseed [12]. Leavened bread was traditionally made
75 with flour, water and a fermenting agent, which was either a fermenting beverage or a fermenting dough,
76 called sourdough. This sourdough was generally initiated from a mixture of flour and water, naturally
77 colonized by lactic acid bacteria (LAB) and yeasts. Sourdough was then either maintained over time or
78 initiated again and again, depending on the craftsman [13, 14]. In the 19th century, the use of
79 yeast starters made of *S. cerevisiae,* often called « baker's yeast », spread as an alternative to sourdough.
80 Nowadays, *S. cerevisiae* industrial starters are more frequently used than sourdoughs, although the latter
81 are gaining interest. A recent study showed that industrial populations of *S. cerevisiae* have followed a
82 different evolutionary path than sourdough populations [10]. Both have been domesticated by humans

which has improved their fermentation performance in a sourdough-mimicking medium. Industrial and sourdough strains of *S. cerevisiae* differ genetically and phenotypically, indicating that sourdough use contributes to the conservation of bread related *S. cerevisiae* lineages [10].

Yeasts, are organisms growing mainly as single cells and with sexual states not enclosed in fruiting bodies, belonging to ascomycete or basidiomycete fungi. To date, more than 40 yeast species have been detected in sourdough [1, 15, 16]. The most frequently encountered species are *Wickerhamomyces anomalus* and *Kazachstania humilis*. Several other species in the genus *Kazachstania* (*Kazachstania barnettii*, *Kazachatania exigua*, *Kazachstania bulderi*, *Kazachstania unispora*) as well as several species in the polyphyletic genus *Pichia* have also been recurrently detected. The factors determining the presence in sourdough of these species are still unknown. A recent large-scale study of 500 sourdoughs from four continents found no effect of geography or factors related to bread making practices such as age of sourdough, storage location, feeding frequency, or grain intake [17]. However, most of the sourdoughs in this study were made by private citizens who probably did not maintain the sourdough microbial community in the same way as professional bakers. To our knowledge, no studies have been conducted to date to investigate the effect of bakers' bread making practices on sourdough yeast community composition.

In France, sourdough breads are made both by bakers and farmers who also grow and mill their own wheat. The number of farmer-bakers has increased in the 2000s with two motives: to grow wheat varieties meeting their needs and to assert their independence from industry [18]. 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, biologists, bio-statisticians, bakers and farmer-bakers to study whether and how bakers and farmer-bakers contribute to the preservation of socio-cultural and fungal species diversity in sourdough microbial community.

Materials and methods

A total of 27 bakers and 12 farmer-bakers participated to the study. They were all making bread with organic flour except five. All of them sent sourdough to the lab for microbiological and metabolic analysis. Among them, 36 described their bread making practices as well.

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

Data on bread making practices were collected 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 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 [19, 20], 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 analyzed. pH and Total Titrable Acidity were measured as described in [20]. Organic acids, alcohol and sugars concentrations (expressed as g/kg of sourdough) were analyzed by liquid chromatography using an HPLC HP 1100 LC system (Agilent

132 technologies, Santa clara, CA, USA) equipped with a refractive index detector (RID Agilent G1382A)
133 and a UV detector (Agilent G1314A). Two different columns were used, a Rezex ROA-organic acids
134 column and a Rezex RPM-monosaccharide column (SDVB – Pb+2 8%, 300x7.8mm, Phenomenex,
135 Torrance, CA, USA). The details of the experiments are described in supplementary information
136 (Method S1).

137 **Yeast species identification**

138 The Internal transcribed spacer 1 (ITS1) ribosomal DNA of each 1216 yeast isolates was amplified by
139 PCR from chromosomal DNA, either by using primers ITS1F and ITS2 [21, 22], or primers NSA3 and
140 58A2R [19, 22]. For isolates unidentified with the ITS1 region alone, DNA was extracted with the
141 MasterPure yeast DNA purification kit (Epicentre, Epibio). PCR reactions targeting partial genes, the
142 D1D2 region of the large subunit of rRNA (LSU), a part of the RNA polymerase II large subunit
143 encoding gene (*RPB1)*, a part of the RNA polymerase II encoding gene (*RPB2)*, a part of the actin
144 encoding gene (*ACT1)* and transcriptional elongation factor (TEF) were performed. To discriminate
145 three specific isolates, PCR on genes *GHD1*, *FSY1*, *URA3*, *DRC1*, *MET2* were performed [23–26]
146 (Table S2). All PCR products were sent to be sequenced with Sanger sequencing (Eurofins, Germany).
147 Species were identified using NCBI [27], YeastIP [28] and a personal database, which was constructed
148 after *ITS1*, *RPB2*, LSU sequencing of all 33 yeast species reportedly found in sourdoughs in the literature
149 [19].

150 **Sourdough DNA extraction, MiSeq sequencing, bioinformatics**

151 The ITS1 region was targeted with the PCR primers ITS1-F (5′- CTTGGTCATTTAGAGGAAGTAA -
152 3′) and ITS2 (5’-GCTGCGTTCTTCATCGATGC-3’).

153 The sequencing run was performed with MiSeq Reagent Kit v3. 2015 [20]. Sequences were analyzed
154 through FROGS “Find Rapidly OTU with Galaxy Solution” [29] and home-made pipelines. Overlapped
155 reads were merged with Flash [30] with a minimum overlap of 10 nucleotides, a maximum overlap of
156 300 nucleotides and a maximum mismatch density of 0.1. Primers were removed with Cutadapt [31]

157 and data were cleaned with Sickle with quality-threshold and length-threshold equal to 20 [32]. Reads
158 were clustered with Swarm (d=3) [33] and chimeras deleted with VSEARCH [34]. Sequences were then
159 filtered on minimum abundance of 0.005% of all sequences. From the OTU abundance table and for
160 each OTU, the taxonomic affiliation using UNITE Version 7.1, Release 2016-11-20 [35], YeastIP [28]
161 and our own databases [19] was obtained by blasting operational taxonomic units (OTUs) representative
162 sequences against each database.

163 **Phenotypic analysis of yeast strains**

164 Fermentation performance of the two most frequently encountered *Kazachstania* species was assessed
165 as described in [10] for *S. cerevisiae*. Fifteen sourdough strains of *K. bulderi* and 16 sourdough strains
166 of *K. humilis* were included in the analysis. *Kazachstania bulderi* strains were coming from sourdoughs
167 B4, B12, B15, B17 B20, B21, while *K. humilis* strains were coming from sourdoughs B2, B5, B6, B7,
168 B10, B17. From one to three strains per sourdough were analyzed in the experiment. In addition, four
169 strains of *K. bulderi* (strain MUCL 38021 isolated from silage in Namur, Belgium, strain MUCL 54694
170 isolated from silage in Erezée, Belgium, strain NRRL Y-27205 and strain CLIB 604 isolated from maize
171 silage in the Netherland) and three strains of *K. humilis* (strain CBS 7754 isolated from food dressing,
172 strain CLIB 1323 isolated from bantu beer, and strain CBS 2664 isolated from alpechin) which were
173 coming from non-sourdough habitats, were added as control to test the effect of habitat of origin. Each
174 strain was phenotyped at least in triplicate leading to a total of 145 fermentations distributed over two
175 blocks. Briefly, fermentations were carried out at 24 °C with constant magnetic stirring (300 rpm) during
176 24 h. CO_2 release was measured by weight loss every 40 min using an automated robotic system [36].
177 At the end of fermentation, population size and cell viability were determined by flow cytometer (C6
178 cytometer, Accuri, BD Biosciences) as described in [37].

179

180 **Data analyses**

181 A multiple correspondence analysis (MCA) and hierarchical clustering (complete linkage clustering
182 method) on principal components based on the first two axes of the MCA were performed using the
183 FactoMineR R package [38].

184 To analyse fungal community, weighted UniFrac distances between sourdough communities were
185 computed from a rooted phylogenetic tree based on the OTU sequences using the R-packages Phyloseq
186 and GUniFrac [39, 40]. Phylogenetic sequences were aligned with Clustalo and phylogenetic trees were
187 built with the parsimony algorithm, with 100 replicates bootstraps, pairwise ktuple-distances with
188 Seaview [41]. The results presented in the main text were obtained using the phylogenetic tree rooted
189 on the OTU identified as *Sporidiobolales species.* Different roots were tested. The roots were chosen
190 among the OTUs that were affiliated to the most distant taxa (*Sporidiobolales species*, *Bullera*
191 *globospora*, *Trichosporon asahii*, *Udeniomyces pyricola*). The tree topology did not change with the
192 chosen root. The tree did not fit the expected phylogeny and, notably, some *Ascomycota* were located
193 among the *Basidiomycota*. However, the dominant sourdough species belonging to the
194 *Saccharomycetaceae* family were clustered in the expected clades or subclades, except that
195 *Kazachstania servazzi* and *Kazachstania unispora* were grouped in a clade closer to *Saccharomyces*
196 species than to other *Kazachstania* species. Using the UniFrac distances matrix, we performed a
197 principal coordinate analysis (PCoA) and clustered sourdough communities using the first two axes of
198 the PCoA, and the complete linkage clustering method (hclust R function). To check the sensitivity of
199 our analysis to this misclassification, we performed the same analyses without the sourdoughs that had
200 one misclassified species representing more than 10% of their reads, *i.e.* sourdoughs B20, B41, B42,
201 and B44 and found the same clustering [40].

202 For each sourdough, the species richness, Chao1, Shannon and Simpson indexes were computed. Chao1
203 was used as an indicator of species richness corrected by the number of OTUs present in the community
204 but not observed. Shannon and Simpson index values were converted to the effective number of species
205 per sourdough. This number was estimated from the Shannon diversity index as $exp^{Shannonindex}$ and

206 from the Simpson diversity index as $\frac{1}{1-Simpsonindex}$ [42, 43]. For probability estimates, the exact 95%
207 confidence intervals were computed using a binomial distribution.

208 To investigate the relationship between β-diversity and differences in bakery practices, we performed a
209 permutational multivariate analysis of variance (PERMANOVA) on the UniFrac distance matrix for
210 each bakery practice variable. We included in the analysis sourdough fungal communities of the 30
211 bakers who had less than 8 missing values among the 29 bread making practices variables and adjusted
212 the p-value using false discovery rate method correction to account for multiple testing [44]. In addition,
213 we performed independence exact Fisher tests between fungal community PCoA groups and each of the
214 bread making practices variables. Multiple testing was accounted for using the false discovery rate
215 method [44].

216 In addition, we tested the link between the baker practice group, the fungal community group or the
217 yeast dominant species and the variation of each quantitative variable (microbial density, pH, TTA,
218 metabolite concentration) with the following mixed effect model: $Y_{ijk} = \mu + \alpha_i + B_j + \varepsilon_{ijk}$ with ε_{ijk}
219 ~N(0,σ^2), where α_i is the effect of the fungal community group i modelled as a fixed effect and B_j is
220 the effect of sourdough j modelled as a random effect and k represents the measurement replicates. For
221 sourdough hydration rate, the variable was arcsin transformed but sourdough effect was not included in
222 the model because no repetition was obtained from any sourdough. The model parameters were
223 estimated using the lmertest R package [45]. To test the fixed effects, we used likelihood ratio tests.
224 Multiple comparisons of means were performed using Tukey tests with the multcomp package. p-values
225 were all adjusted for multiple testing with the FDR method. The geographical structuration was tested
226 with a Mantel test on the UniFrac distances matrix and the geographical distances matrix computed with
227 the package geosphere [46] and ade4 [47].

228 The phenotypic diversity of *K. bulderi* and *K. humilis* strains coming from sourdough and non-
229 sourdough habitats were analyzed. Population size and mortality rate after 27h of fermentation measured
230 by flow cytometer were used as proxies for absolute fitness. The cumulative CO_2 production curve was

231 calculated and the kinetics of CO_2 production rate over time was estimated by successive linear
232 smoothing over five points. Four fermentation parameters were then estimated. The maximum CO_2
233 release (CO_2max, in g/L) was estimated by the maximum of the cumulative CO_2 production curve. The
234 fermentation latency phase time was estimated by the time between inoculation and the beginning of the
235 fermentation calculated as 1g/L of CO_2 release (t1g, in h). The maximum CO_2 production rate (Vmax
236 in g/L/h) was estimated by the maximum of the CO_2 production rate kinetic. The time of the Vmax
237 (tVmax in h) was calculated as the time between inoculation and the Vmax. Hence, the phenotype of
238 each strain was characterized by six quantitative variables called "phenotype variables" below: its
239 population size, its mortality rate and the four fermentation parameters. To determine whether the origin
240 of the strain (sourdough or non-sourdough) had an impact on strain phenotype, each log-transformed
241 quantitative variable was analyzed separately using a mixed linear model as described below. The
242 experimental design was unbalanced between the two blocks with very few non-sourdough strains in
243 one of the two blocks. Therefore, for each phenotype variable, we first estimated the block effect with
244 a subset of 8 strains cultivated in both blocks using a linear model with two fixed effects: the strain and
245 the block. Additive models were used as the interaction terms were not significantly different from zero
246 after adjusting p-values with the Benjamini-Hochberg method. Second, each phenotype variable was
247 corrected for the block effect and analyzed with the mixed effect model: $Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_{ij} +$
248 $Z_k + \epsilon_{ijkl}$where Y_{ijkl}represents the log-transformed phenotype variable corrected for the block effect
249 for the strain *k,* from species *i (i=1,2),* sampled in environment *j (j=1,2)*, observed for replicate *l*. μ
250 represents the mean of the phenotype variable, $\alpha_i$the additive effect of species *i*, $\beta_j$the additive effect
251 of environment *j,* and γ_{ij}their interaction. $Z_k$represents the gaussian random effect of strain *k* with $Z_k \sim$
252 $N(0, \sigma_S^2)$and $\epsilon_{ijkl}$the gaussian residuals with $\epsilon_{ijkl} \sim N(0, \sigma^2)$. For each species *i*, the impact of the
253 environment was quantified using the contrast $\Delta_i = \beta_S + \gamma_{i,S} - \beta_{NS} - \gamma_{i,NS}$ with NS standing for
254 "non-sourdough" and "S" for '"sourdough" and tests ($H_0 = \Delta_i = 0, H_1: \Delta_i \neq 0$) were performed and
255 p-values were adjusted using the Benjamini-Hochberg method. As log-transformed data were analyzed,

the exponential of this contrast can be interpreted as the ratio between the sourdough mean and the non-sourdough mean. Confidence intervals and tests were performed using the doBy R package.

All statistics and plots have been done with R (ggplot2 [48], leaflet package [49], with minor esthetical adjustment with Inkscape). Data and scripts are shared on Zenodo. http://doi.org/10.5281/zenodo.2600170

Results

Two groups of bread making practices

A total of 39 French bakers producing natural sourdough bread and distributed all over France participated to the study (Table S1). The bread making practices of 35 of them were collected through one or several methods: personal interviews (with 12 bakers), focus groups (three groups), observation during bread making workshops (two workshops), and an online/phone survey (36 bakers). The general process of sourdough bread making is presented in Figure 1. We analyzed 28 variables, describing variations of the practices at all steps of the bread making process, from wheat grains to baked bread (Figure S1). Four bakers (B6, B10, B19, B20) who did not provide enough information about their practices were excluded from the multivariate analysis. According to a hierarchical clustering on principal components (HCPC), the 32 other bakers clustered into two groups corresponding to two main types of bread making practices (Figure 2). The first group, hereafter termed “farmer-like” practice group, included six bakers and 11 farmer-bakers using the following practices: low bread production (<500 kg per week, 81% of the bakers of the “farmer-like” group), use of wheat landraces (56%), manual kneading (63%), working at ambient temperature (88%), long fermentation periods (more than 4 hours for 88%), and no use of commercial baker’s yeast (88%). In addition, they tend to make their chief sourdough from dough after kneading (75%). The second group, hereafter called “artisanal-like” practice group consisted of 12 bakers and four farmer-bakers having more intensive practices, characterized by a large bread production (>500 kg per week, 81%), mechanical kneading (100%), use

281 of modern wheat varieties (63%), working at ambient temperature (56%), using *S. cerevisiae* starters in
282 addition to sourdough for bread making or for pastries and buns making (81%). In this second group,
283 bakers tended to make their chief sourdough from a final sourdough.

284 **Composition of sourdough fungal communities**

285 Sourdough is a mix of flour and water naturally fermented by bacteria and yeasts. Sourdough yeast
286 density ranged from 8.1 10^4 to 5.8 10^8 CFU per gram of sourdough, with a mean value of 2.9 10^7 CFU
287 per gram, as commonly found in sourdoughs from all over the world [1, 16, 50, 51]. We isolated 20 to
288 40 yeast strains from each sourdough by picking colonies randomly and identified species using ITS
289 sequence as well as other barcodes when the ITS alone was not able to discriminate between closely
290 related species. Among the 39 collected sourdoughs, one (sourdough B14) did not give any colony in
291 the laboratory, suggesting that his sourdough microbiota was no longer alive. A total of 1216 strains
292 were characterized from the other 38 sourdoughs. In addition, we developed an ITS1 meta-barcoding
293 MiSeq sequencing method on sourdough (see sup M&M). After filtering 5,360,620 raw ITS1 sequences
294 for quality, abundance (0.005%) and chimera, 3,542,801 sequences were further analyzed. Overall, the
295 sequences clustered in 113 OTUs. The number of reads per sourdough ranged from 8421 to 194,557.
296 Therefore, we carried out our analysis on the rarefied matrix. Among all OTUs, 10 were assigned to the
297 order *Triticodae* (especially to the species *Triticum aestivum*), 50 were assigned to a filamentous fungal
298 genus including plant pathogen species such as *Alternaria*, *Aspergillus*, *Fusarium*, or *Gibberella*, while
299 4 OTUs remained unidentified. Among the 40 yeast OTUs, 96% of total reads were assigned to the
300 phylum *Ascomycota*, 87.5% to the order *Saccharomycetales* and 85.7% to the family
301 *Saccharomycetaceae*. Only 4% of the total reads were assigned to the phylum *Basidiomycota*. Overall,
302 three OTUs assigned to the species *Kazachstania humilis*, *Kazachstania bulderi* and *Saccharomyces*
303 *cerevisiae* represented 20.3%, 15.5% and 24.1% respectively of the total number of reads and 28.1%,
304 23.7% and 18.2% respectively of the number of reads identified as yeast species (Figure 3).

305 The non-culture-based and culture-based methods allowed the identification of the same dominant
306 species (defined as a species with an over 50% frequency) for all sourdoughs but five (B09, B20, B22,

307 B25, B41) (Figure 3). In two cases, the discrepancy was explained by the detection of the *Cladosporium*
308 genus at high frequency with metabarcoding while this species could not be isolated in the laboratory
309 (Figure 3). In two other cases, it was explained by a high number of *S. cerevisiae* isolated in the
310 laboratory compared to what was observed using metabarcoding sequencing. In the last case, the
311 identification of *Pichia kudriavzevii* required additional sequencing as its ITS sequence is identical with
312 that of *Candida xylopsoci*. Because metabarcoding allows a deeper characterization of the fungal species
313 diversity with few discrepancy cases, the distribution of fungal species across sourdoughs has been
314 further described using metabarcoding data only. Previous analysis of the same sourdoughs revealed
315 that *Fructilactobacillus sanfranciscensis* was the dominant bacterial species in all analyzed sourdoughs
316 but two, where the dominant species was either *Latilactobacillus curvatus* or *Companilactobacillus*
317 *heilongjiangensis* [20, 52, 53]. Therefore, we decided to study the species composition of the fungal
318 community only.

319 **Fungal species diversity within and between sourdoughs**

320 All sourdoughs but two had a dominant yeast species with a relative abundance over 50% and
321 many species with a lower relative abundance (Figure 3). Within sourdoughs, fungal species richness
322 ranged from 10 to 33, with a 23 median (Table S3). The effective number of species per sourdough
323 calculated from the Shannon diversity index ranged from 1 to 7 (Table S3), with 70% of sourdoughs
324 having an index below two (Table S3). The bread making practice group (artisanal-like/farmer-like) did
325 not influence significantly the level of fungal α-diversity in sourdough (Wilcoxon rank exact test,
326 Wshannon = 156, p-value = 0.16, Wsimpson = 165, p-value = 0.08). Between-sourdough ß-diversity
327 was analyzed using weighted UniFrac distances, computed from a phylogenetic tree built from the
328 distances between OTUs using *Sporidiobolales* species as root (Figure S2). UniFrac distances computed
329 with four differently rooted trees were highly positively correlated (Figure S3). UniFrac distances
330 between sourdoughs ranged from 0.0005 and 0.71, with a median of 0.49 and a mean of 0.52. The
331 clustering of sourdoughs according to their UniFrac distances is shown Figure 5. There was no

332 significant correlation between the UniFrac distances and geographical distances between sourdoughs
333 (Mantel test, P=0.35).

334 We then analyzed specifically the occurrence of yeast species in sourdoughs as yeasts, together
335 with lactic acid bacteria, are the main functional players in a sourdough ecosystem and for bread quality.
336 Over the 40 yeast species detected in the 38 sourdoughs, 12 had a relative abundance over 50% in at
337 least one sourdough, four had a relative abundance between 20% and 50% and 24 had a relative
338 abundance below 10%. All dominant species (relative abundance over 50%) were fermentative yeast
339 species, except in one sourdough that had a C*ladosporium* species. We found all the sourdough yeast
340 genera (*Saccharomyces*, *Kazachstania*, *Pichia*, *Torulaspora* and *Hyphopichia)* commonly reported in
341 the literature except the *Wickerhamomyces* genus that we did not detect in our samples [1, 15, 16].

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

344 *Saccharomyces cerevisiae* was found in 53% of all sourdoughs (95% confidence intervals=36%
345 - 69%) but was dominant (relative abundance over 50%) in only 24% (95% confidence intervals=11%
346 - 40%) (Figure 3). In two cases, *S. cerevisiae* co-occurred with another yeast species at similar relative
347 abundance. In the first case, *S. cerevisiae* was present at a relative abundance of 40% with *Candida sake*
348 at a 41% relative abundance. In the second case, it was found at a relative abundance of 47% with *Pichia*
349 *kudriavzevii* at a relative abundance of 52%. In all the other cases, *S. cerevisiae* had a relative abundance
350 below 21% and was found with other dominant yeast species, such as *Kazachstania australis*,
351 *Kazachstania humilis*, *Saccharomyces uvarum* or *Torulaspora delbrueckii.* This suggests that *S.*
352 *cerevisiae* did not displace other species and can indeed be out-competed by other species in sourdoughs.

353 **Sourdough yeast species mostly belong to the *Kazachstania* genus**

354 *Kazachstania* was the most represented yeast genus over all sourdoughs, when considering both the
355 number of reads over all sourdoughs and the number of detected species. This genus represented 57%
356 of the total number of reads while *Saccharomyces* represented 26% of the total number of reads. In

357 addition, eight species of the *Kazachstania* genus were found in sourdough, while the *Saccharomyces*
358 genus was represented by two species (*S. uvarum* and *S. cerevisiae*) (Figure 3). The *Kazachstania* genus
359 is one of the closest genetically related genus to *Saccharomyces* and contained Crabtree positive yeasts,
360 able to ferment glucose even when oxygen is present if the amount of sugar is sufficient (Hagman &
361 Piskur 2015). *Kazachstania* species dominated in 54% (95% confidence intervals=36%-69%) of
362 sourdoughs while *Saccharomyces* species dominated in 27% only (95% confidence intervals=13%-
363 43%). *Kazachstania humilis*, followed by *K. bulderi,* were the most commonly dominant *Kazachstania*
364 species, and found in respectively 21% (95% confidence intervals=10%-37%) and 15% of sourdoughs
365 (95% confidence intervals=6%-31%) (Figure 3). A recently described *Kazachstania* species,
366 *Kazachstania bozae*, was also identified in five sourdoughs (4.5%-29%) and found dominant in three
367 (1.7%-22%) [64]. Strains of this species were closely related to a strain previously isolated from boza,
368 a Bulgarian fermented drink, as estimated with ITS and LSU (D1D2) barcodes (Source: NCBI,
369 GenBank: KC118125.1 and KX369579.1). In addition, *Kazachstania saulgeensis,* a recently described
370 species [65, 66], was dominant in one sourdough (0.07%-14%). *Kazachstania unispora* and
371 *Kazachstania servazzi* which had previously been detected in sourdough were also found [17, 17, 53,
372 57, 58, 63, 67–71]. Some *Kazachstania* species were detected for the first time as dominant in
373 sourdoughs, whereas they had been previously found in other environments, like soil (*K. australis*) and
374 sauerkraut (*K. barnettii*) [72–74]. None of the previous studies on sourdough have observed as many
375 *Kazachstania* species in sourdough.

376 The composition of sourdough fungal communities was associated with differences in 377 bread making practices

378 We tested whether sourdough fungal community beta diversity could be explained by bread making
379 practices. To do so, we performed univariate PERMANOVA analysis on the 30 bakers with fewer than
380 8 missing values for the 29 bread making practices variables (Table S4). The univariate analysis revealed
381 that the weighted UniFrac distance was structured according to the use of commercial yeast in bakery
382 ($P<0.05$). It also varied significantly with sourdough age, chief sourdough origin (dough, sourdough or

383 both), the quantity of bread produced per week, the milling method (cylinder, millstone, Astrie, Tyrol),
384 the type of wheat variety (ancient, modern or a mix thereof) and the fermentation duration. However,
385 after FDR correction for taking into account multiple testing, none of these variables significantly
386 explained UniFrac distances.

387 In order to understand further the relationship between sourdough fungal community composition
388 and bread making practices, we clustered sourdoughs according to their fungal community composition,
389 on the basis of the PCoA of their weighted UniFrac distances. Then, we tested the link between the
390 fungal community group and the bread making practice group (farmer/artisanal practice group) as well
391 as the link between the fungal community group and each of the different bread making practices (Figure
392 5). Sourdoughs were clustered into three fungal community groups. Group 1 encompassed all
393 sourdoughs (but two) having *Kazachstania* species as dominant species (*K. humilis, K. barnettii, K.*
394 *bulderi, K. saulgeensis, K. bozae)*. Group 2 contained sourdoughs with *Saccharomyces sp.*, *K. servazzi*
395 or *K. unispora* as dominant species. Group 3 harbored sourdoughs with *S. cerevisiae* together with other
396 species such as *Pichia kudriavzevii*, *Candida sake*, or a Dipodascaceae sp. Group 1 sourdoughs were
397 mostly made by bakers having farmer's bread making practices while group 2 and group 3 sourdoughs
398 were mostly made by bakers using artisanal practices (exact Fisher test, P=0.035). The fungal
399 community groups were significantly associated with two specific bread making practice variables: the
400 quantity (in kg) of bread made per week (Exact Fisher test, P=0.001) and the use of commercial yeast
401 (Exact Fisher test, P=0.05). All sourdoughs in group 2 but one were found in bakeries making between
402 500 kg and 1000 kg of bread per week, while groups 1 and 3 sourdoughs originated from bakeries
403 producing very different amounts of bread (ranging from amounts below 250 kg to over 1000 kg). In
404 addition, group 1 sourdoughs were more frequently found in bakeries that do not use commercial yeast
405 while groups 2 and 3 were more frequently found in bakeries using the commercial yeast *S. cerevisiae*
406 (Exact Fisher test, P=0.01). Interestingly, group 1 sourdoughs harbored *S. cerevisiae* either at a relative
407 abundancy below 1% or not at all, while all groups 2 and 3 sourdoughs had *S. cerevisiae* at a relative
408 abundancy over 20%, except in three cases where it was either absent or at a relative abundancy below
409 6%.

410 To test more specifically the link between bread making practices and the distribution of
411 *Kazachstania* species, we analyzed more in-depth group 1 sourdoughs. Within this group, eight
412 sourdoughs had *K. humilis* as dominant species, six had *K. bulderi*, three had *K. bozae* and the remainder
413 had still other *Kazachstania* species. All sourdoughs made with artisanal practices carried *K. humilis* as
414 dominant species or, in one case, the *K. bozae*. By contrast, sourdoughs made with farmer practices had
415 as dominant species *K. bulderi*, *K. australis*, *K. barnettii*, *K. saulgeensis* or *K. bozae* (exact Fisher test,
416 P=0.004).

417

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

420 The composition of fungal community may affect sourdough metabolic content (sugars, acids, alcohols)
421 via fungal strains metabolite consumption and production. Inversely, the presence and concentration of
422 different compounds (sugars, acids, alcohols) may affect differently the fitness depending on the strains
423 and consequently be one of the drivers of fungal community composition. For example, lactic acid
424 bacteria (LAB) are the main producers of acidity in sourdough, but yeasts also produce acetic acid and
425 also indirectly affect acidity through positive or negative interaction with bacteria.

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

435 sugar concentrations. This might be explained by the co-occurrence of bacteria which have their own
436 metabolism and interact by competition and/or cross feeding with sourdough yeasts.

437 We then tested whether the variation of each quantitative variable was associated with the bread
438 making practice groups (farmer-like practices and artisanal practices). There was no significant effect
439 of the bread making practice group except for sourdough hydration that was significantly higher in
440 sourdoughs made using farmer-like practices ($F_{1,94}$=11,69, P<0.001). On average, sourdoughs made with
441 farmer-like practices had 55% water while sourdoughs made with artisanal-like practices had in average
442 49% of water.

443 In addition, we tested whether variations in quantitative variables were associated with the fungal
444 community groups (Table S5). Group 3 microbial community sourdoughs (defined by PCoA clustering
445 on UniFrac distance, see below), which contains *S. cerevisiae* in co-dominance with a second yeast
446 species (*Candida sake*, *Pichia kudriavzevii* or a *Dipodascus* species), had a significantly higher mean
447 pH (mean pH_{group3}=4.2 against pH $_{group1}$=3.8, Tukey Contrasts, P<0.001), lower TTA (mean TTA
448 $_{group3}$=7.7 against TTA $_{group1}$=17.1, Tukey Contrasts, P=0.002), and a higher maltose concentration (mean
449 Maltose $_{group3}$=52.8 mg/gr of sourdough against Maltose $_{group1}$=24.1 mg/gr of sourdough, Tukey
450 Contrasts P=0.002) than group1, having a *Kazachstania* dominant species. Compared to group 2 having
451 in most cases *S. cerevisiae* as dominant species, it also had higher pH (pH $_{group2}$=3.9, Tukey Contrasts,
452 P=0.003), and higher maltose concentration (Maltose $_{group2}$=23.7, Tukey contrast, P=0.003). These data
453 may reflect a lower fermentative activity for group 3 fungal community having two co-dominant species,
454 and/or a negative interaction effect of group 3 fungal community on the activity of lactic acid bacteria
455 (LAB), which are the main producers of sourdough acids. Previous studies on the bacteria content of
456 the same sourdoughs showed that *F. sanfranciscensis* was most generally the dominant species, although
457 *C. heilongjiangensis, L. curvatus* or *Levilactobacillus brevis* were also found as dominant species [20,
458 52, 53]. We found no significant correlation between LAB and yeast densities (r= -0.15, p= 0.45, Figure
459 S4) but the link between fungal and bacterial community might be species and strains dependent.

460 Additional studies on the interactions between fungal and bacterial communities need to be performed
461 to better understand how they may drive sourdough acidity and sugar content.

462 We also analyzed whether the variations of each quantitative variable was associated with the
463 dominant yeast species. We only considered the 26 sourdoughs having either *S. cerevisiae* (9
464 sourdoughs), *K. humilis* (8 sourdoughs), *K. bulderi* (6 sourdoughs) or *K. bozae* (3 sourdoughs) as
465 dominant species, since the other yeast species were found dominant only once. The differences in
466 dominant species was not significantly associated to variation in sourdough sugar, acids or alcohol
467 concentration. However, on average, sourdoughs dominated by *K. bulderi* were more hydrated (63%
468 water content in average) than sourdoughs dominated by *K. humilis*, *K. bozae*, and *S. cerevisiae*, having
469 respectively 49%, 47%, 53 % water content in average ($P<0.001$ for the 3 Tukey Contrasts).
470 *Kazachstania bulderi* was found to be dominant only in sourdoughs made using farmer-like practices, a
471 bread making practice group that was also found to be associated with more hydrated sourdoughs.
472 Additional experiments should be carried out to test whether this species has indeed a better fitness in
473 more hydrated sourdoughs or whether its presence in more hydrated sourdoughs is related to covariation
474 with other farmer practices.

475 In conclusion, no clear evidence was found of the impact of bread making practices or of the
476 dominant yeast species on the metabolic composition of sourdough. On the other hand, our results
477 showed metabolic differences between sourdoughs having one or two co-dominant yeast species.

478 **Phenotypic signatures of domestication**

479 A previous analysis on *S. cerevisiae* revealed that sourdough strains had higher average fitness and
480 fermentation performance than strains from other environments in a sourdough-mimicking medium
481 [10]. Here, we investigated whether evidence of a domestication syndrome could also be found in *K.*
482 *humilis* and *K. bulderi*, the two *Kazachstania* species most commonly found in French sourdoughs. We
483 tested whether fitness (log of population size and mortality at the end of fermentation) and fermentation

484 performance (CO2max, Vmax, t1g, tVmax) differed between sourdough strains and strains from
485 elsewhere.

486 A principal component analysis of 38 strains of *K. bulderi* and *K. humilis*, based on quantitative variation
487 in the six phenotypic variables described below was carried out. The first two axis explained 80.5% of
488 the variation and clearly separated strains by species (Figure 6). The *K. bulderi* strains were located at
489 the right of the PCA and were characterized by high population size and low mortality at the end of
490 fermentation, while the *K. humilis* strains were located at the left and were characterized by a rapid onset
491 of fermentation (t1g), high maximum fermentation rate (Vmax), and a short time to reach Vmax
492 (tVmax). Non-sourdough strains of *K. humilis* were located outside the cloud of sourdough strains while
493 non-sourdough strains of *K. bulderi* were distributed within and outside the cloud of sourdough strains.

494 Statistical comparisons of sourdough and non-sourdough strains of *K. bulderi* and *K. humilis* for each
495 phenotypic variable revealed phenotypic divergence for *K. humilis* but not for *K. bulderi*. While the *K.*
496 *bulderi* sourdough strains did not ferment significantly faster than the non-sourdough strains, the *K.*
497 *humilis* sourdough strains showed significantly higher Vmax, lower t1g, and lower tVmax than the non-
498 sourdough *K. humilis* strains (Figure 7, Table S6). On average, they started fermentation two hours
499 before the others and reached Vmax three hours before the others. In addition, their Vmax were on
500 average 34% higher than the others.

501 Discussion

502 Sourdough microbial diversity has been intensively studied worldwide. Despite a cultural and historical
503 interest on bread in France, French sourdough fungal diversity was only partly characterized before this
504 study [19, 53, 71]. A recent large-scale (> 500 starters) study of sourdough microbial diversity revealed
505 the fungal diversity that can be detected over the globe across home-made sourdoughs [17]. All the
506 yeast species detected at a relative abundance over 1% in this international collection of sourdoughs
507 were detected in French sourdoughs except the species *Wickerhamomyces anomalus*, *Pichia*
508 *membranifaciens*, *Naumovozyma castellii* and *Saccharomyces bayanus*. Inversely, French baker's

509 sourdough harbored some yeast species that were never found elsewhere, such as *K. bozae*, *K. australis*,
510 *K. saulgeensis* [65].

511 Beyond the genus of the baker yeast species, *S. cerevisiae*, the most represented yeast genus in French
512 sourdough was *Kazachstania.* Eight *Kazachstania* species were found as dominant yeast species in at
513 least one French sourdough: *K. humilis*, *K. bulderi*, *K. barnettii*, *K. unispora*, *K. servazzii*, *K. bozae*, *K.*
514 *australis* and *K. saulgeensis.* Three *Kazachstania* species (*K. exigua, K. lodderae, K. naganishi*) already
515 reported in sourdough were not found in our collection of French sourdough. *Kazachstania lodderae*
516 and *K. naganishi* are rarely found in sourdough. By contrast, *K. exigua* is a frequently cited sourdough
517 species in the literature. This species has been previously found in France, Finland [77], Italy [78, 79],
518 Denmark [80], Ethiopia [81], USA [17] and is the first species to have been isolated from a sourdough
519 (in San Francisco, [29]). However, its taxonomic characterization may have been hampered by the fact
520 that it probably originated from hybridization between unknown yeast species [82]. To date, the genus
521 *Kazachstania* is composed of more than 40 species, of which 11 are present in sourdough. It is possible
522 that an adaptive radiation linked to the adaptation to different sourdoughs or to different anthropized
523 niches has taken place, as it has been observed for example in cichlids during their adaptation to different
524 lakes or in Penicilium domesticated fungi. Indeed, five of the *Kazachstania* species present in the
525 sourdough have so far only been detected in human-related niches. These are *K. saulgeensis*, *K.*
526 *barnettii*, *K. bozae*, *K. bulderi* and *K. humilis*. These species are genetically closer to each other than
527 they are to *K. australis*, *K. servazzii* and *K. unispora* which have also been found in nature and are
528 grouped in another part of the *Kazachstania* pylogenetic tree. Genomic analysis of these species would
529 shed light on their evolution and the genetic changes that would have been selected during potential
530 domestication. So far, eight of the 11 *Kazachstania* species found in the sourdough have at least one
531 genome assembly available in public databases, including *K. saulgeensis* and *K. barnettii* [44, 45], and
532 the assemblies of *K. bulderi* and *K. humilis*, which were recently published in public databases (Bio
533 project: PRJEB44438). The genomic and phenomic analysis of the large collection of *Kazachstania*
534 strains obtained by our study, together with the world collection of *Kazachstania* strains, may shed light
535 on the radiation and domestication processes of these species.

We found that yeast community composition was partly related with bread making practices. Bread making practice divergence also led to different phenotypic signatures. Strains of the *K. humilis* species, which was typical of sourdough made by artisanal-like practices, had higher fermentation rate while strains of *K. bulderi* which was typical of sourdough made by farmer-like bread making practices had not. The species *K. humilis* has been found in many countries, viz. Austria, Canada, China, Denmark, Ethiopia, Finland, Germany, Greece, Italy, Morocco, the Netherland, Spain, UK, USA, and France [1, 15–17, 19, 53, 55–60]. It is also the most frequently encountered *Kazachstania* species in sourdoughs around the world. This species is therefore frequently found in bakeries, where short fermentations are often favored. This may explain why sourdough strains of *K. humilis* seem to have been selected for increased fermentation rate. Increased fermentation rate was also found in bakery strains of *S. cerevisiae* when compared to non-bakery strains.

In contrast, we did not find evidence for improved fermentation performance in *K. bulderi*, which was the third most represented species in French sourdough. *Kazachstania bulderi* was found in bakeries with farmer-like practices. These bakeries often bake bread once or twice a week and store their sourdough for several days. They also often use long fermentation and thus may not have selected an increased fermentation rate. Farmer-bakers typically store their sourdough for several days and therefore make a lower number of backslopping. In addition, they make bread with longer fermentation times than artisanal bakers. It is therefore possible that they did not select to accelerate the speed of fermentation and instead let natural selection in the sourdough environment act alone. Alternatively, the lack of phenotypic divergence between sourdough and non-sourdough strains may reflect the limitation of our sampling. *Kazachstania bulderi* has been reported for the first time in anaerobic maize silage in the Netherlands and in fermented liquid feed for piglets [61, 62], more recently in French, Belgium and Spain sourdoughs [19, 53, 57, 63] but to our knowledge was never found in wild environment. Here, we compare sourdough strains with strains coming from silage and animal feed. It is unknown whether silage and animal feed strains are wild strains or feral strains that have escaped from other domesticated environments. This may explain why we did not detect any phenotypic divergence between sourdough and non-sourdough strains of *K. bulderi*. Other than fermentation phenotypes, there was no evidence of

563 fitness differences between sourdough and non-sourdough strains of *K. humilis* and *K. bulderi* in
564 sourdough mimicking media. Additional experiments in real dough should be performed to further test
565 the effect of natural selection in this environment.

566 Other evolutionary process than selection could also explain the distribution of yeast species across
567 sourdoughs. Interviews with the bakers working with sourdough hosting *K. bulderi* and *K. bozae*
568 suggested the role of dispersion of these species in French sourdoughs. Indeed, these bakers have been
569 connected over the years either through seed exchanges, sourdough mixing or gifts, bread making
570 training in common or working in one another's bakery. Some yeast species have been found in the
571 bakery house environment and baker's hands and may thus be dispersed through baker's tools or baker's
572 travels [58, 67, 75, 76]. However, it is still unclear whether wheat seeds and flour are a source of
573 sourdough yeasts.

574 To our knowledge, this is the first evidence of the influence of artisanal practices on taxonomic
575 diversity in microbial communities. On the other hand, several studies have shown that making
576 fermented products could lead to the selection of divergent phenotypes and genotypes. This is the case
577 of sourdough and industrial populations of the baker's yeast *S. cerevisiae* that diverge from each other
578 and have a better fermentation performance than non-baker's strains. As for beer populations, sourdough
579 populations have acquired a better capacity to assimilate maltose, linked at least in part to an increase
580 in the number of copies of the genes involved in the assimilation of maltose. Several studies on wine
581 populations of *S. cerevisiae*, *Torulaspora delbrueckii* and *Lachancea thermotolerans* also showed that
582 wine populations are genetically differentiated from strains from other environments and present
583 beneficial phenotypes in grape must and for wine quality. The analysis of the filamentous cheese fungi
584 *P. roqueforti* and *P. camemberti* also revealed genetically differentiated cheese populations.
585 Interestingly, different genetic groups associated with different cheese making practices were found.
586 Strains of the blue cheese fungus, *P. roqueforti*, isolated from Roquefort cheeses were more diverse and
587 were genetically and phenotypically different than strains used to make other blue cheeses [83, 84, 85].
588 Two varieties of the soft-cheese making fungus, *P. camemberti*, with different phenotypic features, were

associated with different kinds of cheese (Camembert and Brie). All together these studies show that the diversity of practices used to make fermented products allows to maintain genetic, phenotypic and taxonomic diversity.

However, fungal domestication also involved strong bottlenecks. For example, the low level of genetic diversity found in blue-cheese *P. roqueforti* strains and soft-cheese *P. camemberti* strains revealed the risk of diversity erosion in organisms used for fermented product making [6, 7, 84]. This risk is accentuated by fertility depression among fungal domesticated strains [86]. This risk is also associated with the massive use of few industrial strains or the need to standardize products or meet the specifications of industrial production or protected designation of origin (PDO) [6]. Here, we show that despite the recurrent use of *S. cerevisiae* as industrial starter species in bakeries and homes, and the occurrence of this species in a wide range of habitats such as soil, trees, and humans [9, 54], this species does not appear to have overwhelmingly colonized French traditional sourdoughs (Figure 3, Figure 4). This result confirmed a recent analysis which revealed that *S. cerevisiae* sourdough strains had a different evolutionary history from industrial strains [10]. The dynamic of microbial species colonization and invasion in food environment remains largely unknown. Additional experiments at the level of microbial community will shed light on the dynamic of microbial community establishment in food production and on the ability of industrial strains to invade food microbial community.

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 neighboring genus *Kazachstania* were detected at high frequency, revealing a major role for this mostly unknown genus in the study of fungal domestication and in bread making. Therefore, our results highlight the necessity of maintaining socio-cultural diversity to maintain

microbial diversity in food systems. These findings could not have been revealed without the collaboration of bakers and scientists, showing the importance of participatory research projects to gain new insights 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 INRAE. We thank all the bakers and farmer-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* ; **0**.

2. Gibbons JG, Rinker DC. The genomics of microbial domestication in the fermented food environment. *Current Opinion in Genetics & Development* 2015; **35**: 1–8.

3. Steensels J, Gallone B, Voordeckers K, Verstrepen KJ. Domestication of Industrial Microbes. *Current Biology* 2019; **29**: R381–R393.

4. Gibbons JG, Salichos L, Slot JC, Rinker DC, McGary KL, King JG, et al. The evolutionary imprint of domestication on genome variation and function of the filamentous fungus Aspergillus oryzae. *Curr Biol* 2012; **22**: 1403–1409.

5. Watarai N, Yamamoto N, Sawada K, Yamada T. Evolution of Aspergillus oryzae before and after domestication inferred by large-scale comparative genomic analysis. *DNA Research* 2019; **26**: 465–472.

6. Ropars J, Didiot E, Rodríguez de la Vega RC, Bennetot B, Coton M, Poirier E, et al. Domestication of the Emblematic White Cheese-Making Fungus Penicillium camemberti and Its

639 Diversification into Two Varieties. *Curr Biol* 2020; **30**: 4441-4453.e4.
640 7. Ropars J, López-Villavicencio M, Snirc A, Lacoste S, Giraud T. Blue cheese-making has
641 shaped the population genetic structure of the mould Penicillium roqueforti. *PLOS ONE* 2017; **12**:
642 e0171387.
643 8. Legras J-L, Galeote V, Bigey F, Camarasa C, Marsit S, Nidelet T, et al. Adaptation of *S.*
644 *cerevisiae* to Fermented Food Environments Reveals Remarkable Genome Plasticity and the
645 Footprints of Domestication. *Mol Biol Evol* 2018; **35**: 1712–1727.
646 9. Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, et al. Genome evolution
647 across 1,011 *Saccharomyces* cerevisiae isolates. *Nature* 2018; **556**: 339–344.
648 10. Bigey F, Segond D, Friedrich A, Guezenec S, Bourgais A, Huyghe L, et al. Evidence for Two
649 Main Domestication Trajectories in *Saccharomyces* cerevisiae Linked to Distinct Bread making
650 Processes. *Current Biology* 2020; S0960982220316912.
651 11. Arranz-Otaegui A, Carretero LG, Ramsey MN, Fuller DQ, Richter T. Archaeobotanical
652 evidence reveals the origins of bread 14,400 years ago in northeastern Jordan. *PNAS* 2018; **115**: 7925–
653 7930.
654 12. Samuel D. Bread in archaeology. *Civilisations Revue internationale d'anthropologie et de*
655 *sciences humaines* 2002; 27–36.
656 13. Roussel P, Chiron H. Les pains français: évolution, qualité, production. 2005. Maé-Erti.
657 14. Decock P, Cappelle S. Bread technology and sourdough technology. *Trends in Food Science*
658 *& Technology* 2005; **16**: 113–120.
659 15. Van Kerrebroeck S, Maes D, De Vuyst L. Sourdoughs as a function of their species diversity
660 and process conditions, a meta-analysis. *Trends in Food Science & Technology* 2017; **68**: 152–159.
661 16. Arora K, Ameur H, Polo A, Di Cagno R, Rizzello CG, Gobbetti M. Thirty years of knowledge
662 on sourdough fermentation: A systematic review. *Trends in Food Science & Technology* 2021; **108**:
663 71–83.
664 17. Landis EA, Oliverio AM, McKenney EA, Nichols LM, Kfoury N, Biango-Daniels M, et al.
665 The diversity and function of sourdough starter microbiomes. *eLife* 2021; **10**: e61644.

666 18. Demeulenaere É, Rivière P, Hyacinthe A, Baltassat R, Baltazar S, Gascuel J-S, et al. La
667 sélection participative à l'épreuve du changement d'échelle. À propos d'une collaboration entre
668 paysans sélectionneurs et généticiens de terrain. *Natures Sciences Societes* 2017; **Vol. 25**: 336–346.
669 19. Urien C, Legrand J, Montalent P, Casaregola S, Sicard D. Fungal Species Diversity in French
670 Bread Sourdoughs Made of Organic Wheat Flour. *Front Microbiol* 2019; **10**.
671 20. Michel E, Monfort C, Deffrasnes M, Guezenec S, Lhomme E, Barret M, et al.
672 Characterization of relative abundance of lactic acid bacteria species in French organic sourdough by
673 cultural, qPCR and MiSeq high-throughput sequencing methods. *International Journal of Food*
674 *Microbiology* 2016; **239**: 35–43.
675 21. White TJ, Bruns T, Lee S, Taylor J. 38 - AMPLIFICATION AND DIRECT SEQUENCING
676 OF FUNGAL RIBOSOMAL RNA GENES FOR PHYLOGENETICS. In: Innis MA, Gelfand DH,
677 Sninsky JJ, White TJ (eds). *PCR Protocols*. 1990. Academic Press, San Diego, pp 315–322.
678 22. Martin KJ, Rygiewicz PT. Fungal-specific PCR primers developed for analysis of the ITS
679 region of environmental DNA extracts. *BMC Microbiology* 2005; **5**: 28.
680 23. Peris D, Sylvester K, Libkind D, Gonçalves P, Sampaio J, Alexander WG, et al. Population
681 structure and reticulate evolution of *Saccharomyces* eubayanus and its lager-brewing hybrids. *Mol*
682 *Ecol* 2014; **23**: 2031–2045.
683 24. Nguyen H-V, Gaillardin C. Evolutionary relationships between the former species
684 *Saccharomyces* uvarum and the hybrids *Saccharomyces* bayanus and *Saccharomyces* pastorianus;
685 reinstatement of *Saccharomyces* uvarum (Beijerinck) as a distinct species. *FEMS Yeast Res* 2005; **5**:
686 471–483.
687 25. Libkind D, Hittinger C, of the … V-E. Microbe domestication and the identification of the
688 wild genetic stock of lager-brewing yeast. *Proceedings of the …* 2011.
689 26. Sampaio JP, Gonçalves P. Natural populations of *Saccharomyces* kudriavzevii in Portugal are
690 associated with oak bark and are sympatric with *S. cerevisiae* and S. paradoxus. *Appl Environ*
691 *Microbiol* 2008; **74**: 2144–2152.
692 27. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*

693 2016; **44**: D7–D19.

694 28. Weiss S, Samson F, Navarro D, Casaregola S. YeastIP: a database for identification and
695 phylogeny of Saccharomycotina yeasts. *FEMS Yeast Research* 2013; **13**: 117–125.

696 29. Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, et al. FROGS: Find,
697 Rapidly, OTUs with Galaxy Solution. *Bioinformatics* 2018; **34**: 1287–1294.

698 30. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome
699 assemblies. *Bioinformatics* 2011; **27**: 2957–2963.

700 31. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
701 *EMBnet.journal* 2011; **17**: 10–12.

702 32. Sickle 1.33 – Windowed Adaptive Trimming for Fastq files using Quality – My Biosoftware –
703 Bioinformatics Softwares Blog.

704 33. Mahé F, Rognes T, Quince C, Vargas C de, Dunthorn M. Swarm: robust and fast clustering
705 method for amplicon-based studies. *PeerJ* 2014; **2**: e593.

706 34. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool
707 for metagenomics. *PeerJ* 2016; **4**: e2584.

708 35. Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, et al. Towards a
709 unified paradigm for sequence-based identification of fungi. *Molecular Ecology* 2013; **22**: 5271–5277.

710 36. Bloem A, Rollero S, Seguinot P, Crépin L, Perez M, Picou C, et al. Workflow Based on the
711 Combination of Isotopic Tracer Experiments to Investigate Microbial Metabolism of Multiple Nutrient
712 Sources. *J Vis Exp* 2018.

713 37. Delobel P, Pradal M, Blondin B, Tesniere C. A 'fragile cell' sub-population revealed during
714 cytometric assessment of *Saccharomyces* cerevisiae viability in lipid-limited alcoholic fermentation.
715 *Lett Appl Microbiol* 2012; **55**: 338–344.

716 38. Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. *Journal of*
717 *Statistical Software* 2008; **25**: 1–18.

718 39. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and
719 Graphics of Microbiome Census Data. *PLOS ONE* 2013; **8**: e61217.

720 40. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial
721 communities. *Appl Environ Microbiol* 2005; **71**: 8228–8235.
722 41. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface
723 for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010; **27**: 221–224.
724 42. Chao A, Chiu C-H, Jost L. Phylogenetic diversity measures based on Hill numbers. *Philos*
725 *Trans R Soc Lond B Biol Sci* 2010; **365**: 3599–3609.
726 43. Jost L. Partitioning diversity into independent alpha and beta components. *Ecology* 2007; **88**:
727 2427–2439.
728 44. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful
729 Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*
730 1995; **57**: 289–300.
731 45. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed
732 Effects Models. *Journal of Statistical Software* 2017; **82**: 1–26.
733 46. Hijmans RJ, Karney (GeographicLib) C, Williams E, Vennes C. geosphere: Spherical
734 Trigonometry. 2021.
735 47. Dray S, Dufour A-B. The ade4 Package: Implementing the Duality Diagram for Ecologists.
736 *Journal of Statistical Software* 2007; **22**: 1–20.
737 48. Pedersen HW Danielle Navarro, and Thomas Lin. Welcome | ggplot2.
738 49. Cheng J, Karambelkar B, Xie Y. Leaflet: Create interactive web maps with the
739 javascript'leaflet'library. *R package version* 2018; **2**.
740 50. Clément H, Prost C, Chiron H, Ducasse MB, Della Valle G, Courcoux P, et al. The effect of
741 organic wheat flour by-products on sourdough performances assessed by a multi-criteria approach.
742 *Food Res Int* 2018; **106**: 974–981.
743 51. De Vuyst L, Van Kerrebroeck S, Leroy F. Microbial Ecology and Process Technology of
744 Sourdough Fermentation. *Adv Appl Microbiol* 2017; **100**: 49–160.
745 52. Lhomme E, Orain S, Courcoux P, Onno B, Dousset X. The predominance of Lactobacillus
746 sanfranciscensis in French organic sourdoughs and its impact on related bread characteristics.

747 *International Journal of Food Microbiology* 2015; **213**: 40–48.
748 53. Lhomme E, Urien C, Legrand J, Dousset X, Onno B, Sicard D. Sourdough microbial
749 community dynamics: An analysis during French organic bread making processes. *Food Microbiol*
750 2016; **53**: 41–50.
751 54. Goddard MR, Greig D. *Saccharomyces* cerevisiae: a nomadic yeast with no niche? *FEMS*
752 *Yeast Res* 2015; **15**: fov009.
753 55. Valmorri S, Tofalo R, Settanni L, Corsetti A, Suzzi G. Yeast microbiota associated with
754 spontaneous sourdough fermentations in the production of traditional wheat sourdough breads of the
755 Abruzzo region (Italy). *Antonie Van Leeuwenhoek* 2010; **97**: 119–129.
756 56. Wang X, Du H, Zhang Y, Xu Y. Environmental Microbiota Drives Microbial Succession and
757 Metabolic Profiles during Chinese Liquor Fermentation. *Appl Environ Microbiol* 2018; **84**: e02369-
758 17.
759 57. Chiva R, Celador-Lera L, Uña JA, Jiménez-López A, Espinosa-Alcantud M, Mateos-
760 Horganero E, et al. Yeast Biodiversity in Fermented Doughs and Raw Cereal Matrices and the Study
761 of Technological Traits of Selected Strains Isolated in Spain. *Microorganisms* 2020; **9**: 47.
762 58. Reese AT, Madden AA, Joossens M, Lacaze G, Dunn RR. Influences of Ingredients and
763 Bakers on the Bacteria and Fungi in Sourdough Starters and Bread. *mSphere* 2020; **5**.
764 59. Syrokou MK, Themeli C, Paramithiotis S, Mataragas M, Bosnea L, Argyri AA, et al.
765 Microbial Ecology of Greek Wheat Sourdoughs, Identified by a Culture-Dependent and a Culture-
766 Independent Approach. *Foods* 2020; **9**: 1603.
767 60. Sugihara TF, Kline L, Miller MW. Microorganisms of the San Francisco sour dough bread
768 process. I. Yeasts responsible for the leavening action. *Appl Microbiol* 1971; **21**: 456–458.
769 61. Middelhoven WJ, Kurtzman CP, Vaughan-Martini A. *Saccharomyces* bulderi sp. nov., a yeast
770 that ferments gluconolactone. *Antonie Van Leeuwenhoek* 2000; **77**: 223–228.
771 62. Gori K, Bjørklund MK, Canibe N, Pedersen AØ, Jespersen L. Occurrence and identification
772 of yeast species in fermented liquid feed for piglets. *Microb Ecol* 2011; **61**: 146–153.
773 63. Comasio A, Verce M, Van Kerrebroeck S, De Vuyst L. Diverse Microbial Composition of

774 Sourdoughs From Different Origins. *Front Microbiol* 2020; **11**: 1212.
775 64. Gouliamova D, Dimitrov R. *KAZACHSTANIA* CHRYSOLINAE AND *KAZACHSTANIA*
776 BOZAE TWO NEW YEAST SPECIES OF THE GENUS *KAZACHSTANIA*. TRANSFER OF FOUR
777 *KAZACHSTANIA* SPECIES TO GRIGOROVIA GEN. NOV. AS NEW COMBINATIONS. 2020.
778 65. Jacques N, Sarilar V, Urien C, Lopes MR, Morais CG, Uetanabaro APT, et al. Three novel
779 ascomycetous yeast species of the *Kazachstania* clade, *Kazachstania* saulgeensis sp nov.,
780 *Kazachstania* serrabonitensis sp nov and *Kazachstania* australis sp nov Reassignment of Candida
781 humilis to *Kazachstania* humilis f.a. comb. nov and Candida pseudohumilis to *Kazachstania*
782 pseudohumilis f.a. comb. nov. *International Journal of Systematic and Evolutionary Microbiology*
783 2016; **66**: 5192--5200.
784 66. Sarilar V, Sterck L, Matsumoto S, Jacques N, Neuvéglise C, Tinsley CR, et al. Genome
785 sequence of the type strain CLIB 1764T (=CBS 14374T) of the yeast species *Kazachstania*
786 saulgeensis isolated from French organic sourdough. *Genomics Data* 2017; **13**: 41–43.
787 67. Minervini F, Lattanzi A, De Angelis M, Di Cagno R, Gobbetti M. Influence of Artisan
788 Bakery- or Laboratory-Propagated Sourdoughs on the Diversity of Lactic Acid Bacterium and Yeast
789 Microbiotas. *Appl Environ Microbiol* 2012; **78**: 5328–5340.
790 68. Salovaara H, Savolainen J. Yeast type isolated from Finnish sour rye dough starters. *Acta*
791 *Alimentaria Polonica* 1984; **10**.
792 69. Decimo M, Quattrini M, Ricci G, Fortina MG, Brasca M, Silvetti T, et al. Evaluation of
793 microbial consortia and chemical changes in spontaneous maize bran fermentation. *AMB Express*
794 2017; **7**: 205.
795 70. Fraberger V, Unger C, Kummer C, Domig KJ. Insights into microbial diversity of traditional
796 Austrian sourdough. *LWT* 2020; **127**: 109358.
797 71. Lhomme E, Lattanzi A, Dousset X, Minervini F, De Angelis M, Lacaze G, et al. Lactic acid
798 bacterium and yeast microbiotas of sixteen French traditional sourdoughs. *Int J Food Microbiol* 2015;
799 **215**: 161–170.
800 72. 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.

73. 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 Zygotorulaspora. *FEMS yeast research* 2003; **4**: 233–245.

74. Boynton PJ, Greig D. The ecology and evolution of non-domesticated *Saccharomyces* species. *Yeast* 2014; **31**: 449–462.

75. 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

76. Minervini F, Lattanzi A, Dinardo FR, De Angelis M, Gobbetti M. Wheat endophytic lactobacilli drive the microbial and biochemical features of sourdoughs. *Food Microbiol* 2018; **70**: 162–171.

77. Infantes M, Schmidt JL. Characterisation of the yeast flora of natural sourdoughs located in various French areas. Sciences des Aliments (France) [Internet]. 1992, available on: https://agris.fao.org/agris-search/search.do?recordID=FR9204966search/search.do?recordID=FR9204966

78. Pulvirenti A, Solieri L, Gullo M, Vero LD, Giudici P. Occurrence and dominance of yeast species in sourdough. *Letters in Applied Microbiology* 2004; **38**: 113-7.

79. Minervini F, Lattanzi A, De Angelis M, Di Cagno R, Gobbetti M. Influence of Artisan Bakery- or Laboratory-Propagated Sourdoughs on the Diversity of Lactic Acid Bacterium and Yeast Microbiotas. *Appl Environ Microbiol* 2012; **78**: 5328-40.

80. Nuobariene L, Hansen ÅS, Arneborg N. Isolation and identification of phytase-active yeasts from sourdoughs. LWT - *Food Science and Technology* 2012; **48**:190-6.

81. Desiye A, Abegaz K. Isolation, characterization and identification of lactic acid bacteria and yeast involved in fermentation of Teff (EragrostisTef) Batter. 2013; /paper/Isolation%2C-characterization-and-identification-of-Desiye Abegaz/3373ba786cbed5f07d2c6330a5362ea56198adc2

82. Bon E, Neuvéglise C, Lépingle A, Wincker P, Artiguenave F, Gaillardin C, et al. Genomic Exploration of the Hemiascomycetous Yeasts: 6. *Saccharomyces* exiguus. *FEBS Letters* 2000; **487**: 42-6.

83. Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, et al. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. *Nat Commun* 2014; **5**: 2876.

84. Dumas E, Feurtey A, Rodríguez de la Vega RC, Le Prieur S, Snirc A, Coton M, et al. Independent domestication events in the blue-cheese fungus Penicillium roqueforti. *Molecular Ecology* 2020; **29**: 2639-60.

85. Ropars J, Rodríguez de la Vega RC, López-Villavicencio M, Gouzy J, Sallet E, Dumas É, et al. Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi. *Current Biology* 2015; **25**: 2562-9.

86. Ropars J, Lo Y, Dumas E, Snirc A, Begerow D, Rollnik T, et al. Fertility depression among cheese-making Penicillium roqueforti strains suggests degeneration during domestication. *Evolution* 2016; **70**: 2099–2109.

845 **Figure Legends**

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

859

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

862 A) Representation of bakers. Each point represents a bakery. The purple area on the left brings together
863 baker with "artisanal" practices and the light blue area on the right the bakers with "farmer" practices.
864 The dot's colors indicate the PCoA cluster of the sourdough fungal community (see Figure 5). Black
865 dots for group 1, empty dot for group 2, grey for group 3. The fungal community of the sourdough of
866 baker 14 was not studied. B) Representation of the 20 first categories that contributed the most to the
867 MCA axes. The category, which corresponds to a class of a variable, is written next to the triangle. C)
868 Distribution of each variable for each bread making practice group. Only variables that mostly explained
869 differences between bread making practice groups are shown: use of commercial yeast, kneading
870 method, chief origin, kg of bread production per week, number of bread making per week, percent of

871 water in dough, number of back-sloppings before making bread, water origin, sourdough age and flour
872 percentage in dough. The categories of each of these variables are indicated on the right.
873
874 **Figure 3.** Yeast species diversity was analyzed for 38 out of the 39 sourdoughs with both cultural and
875 metabarcoding methods. Left: species were identified by traditional microbial isolation and
876 identification using ITS sequencing. Right: species were identified using ITS1 metabarcoding. The
877 three most frequently-encountered species are shown in contrasting colors surrounded by black (blue:
878 *Saccharomyces cerevisiae*, red: *Kazachstania bulderi*, yellow: *Kazachstania humilis*). The bread
879 making practice of the baker who supplied the sourdough is indicated on the left ("artisanal" in purple,
880 "farmer" in light blue).
881
882 **Figure 4.** Yeast species composition of 38 French sourdoughs. Each bar represents one sourdough and
883 is placed on the map where the sourdough was collected. Sourdoughs coming from bakers with a
884 "farmer" practice are bordered in light blue and labelled with an "F" while sourdoughs coming from
885 bakers with "artisanal" practice are bordered by dark blue and labelled with and "A". Each species is
886 represented by a different color. The relative abondance of each species, estimated by metabarcoding
887 analysis, is represented by the area of each bar, as shown at the top left.
888
889 **Figure 5.** Clustering of sourdough fungal communities based on weighted UniFrac distances and
890 association with baking practices. The weighted UniFrac distance between communities takes into
891 account the relative abundance of the observed species as well as their phylogenetic relationships. In
892 all panels, the clustering of sourdoughs fungal communities based on their Unifrac distances is shown
893 on a tree on the left and on a Principal Component Analysis (PcoA) in the center. Panel A shows the
894 three groups of fungal communities formed based on their weighted UniFrac distances. Panels B through
895 E highlight the distribution of a principal variable in these three groups of fungal communities. The
896 distribution is shown by coloring the different levels of the variable on the tree and on the PCoA but
897 also as a barplot on the right side of the panel. Panel B shows the distribution of the two types of bread

making practices, panel C the dominant yeast species, panel D the use of a commercial yeast starter, panel E the amount of bread made each week. Group 1 contains sourdough fungal communities whose dominant species is a *Kazachstania* species. Group 2 contains sourdough fungal communities whose dominant species is *Saccharomyces cerevisiae*. Group 3 contains sourdough fungal communities in which two yeast species co-occurred, one being *S. cerevisiae*. The yeast genus in panel C is abbreviated with, *C.* for *Candida*, *D.* for *Dipodascaceae*, *K.* for *Kazachstania*, *S.* for *Saccharomyces*, *T.* for *Torulaspora.*

Figure 6. Principal component analysis of 37 *K. humilis* and *K. bulderi* strains based on the quantitative variation of maximum CO2 production (CO2max), fermentation latency phase (t1g), maximum CO2 production rate (Vmax), time to reach the maximum production rate (tVmax), log of population size and mortality at the end of the fermentation. The correlations between variables are presented on the left while the figure on the right shows the projection of strains on the first two axes representing 70.64% of the variation. The strains are colored according to their habitat of origin. Sourdough strains are indicated in blue and non-sourdough strains in red. Their species is indicated by symbol. *Kazachstania humilis* strains are indicated by a circle and *Kazachstania bulderi* strains by a triangle.

Figure 7. Ratio between the sourdough strains mean and the non-sourdough strains mean values of each quantitative variable measuring fermentation performance: time to reach the maximum production rate (tVmax), maximum CO2 production rate (Vmax), maximum CO2 production (CO2max), log of population size after 27 hours of fermentation (Cellt27), cell mortality after 27 hours of fermentation (Mortality), fermentation latency phase (t1g). Confidence intervals are indicated by bars. Mean ratio and confidence intervals are shown in red for *Kazachstania bulderi* and blue for *Kazachtania humilis.*

Figure 1

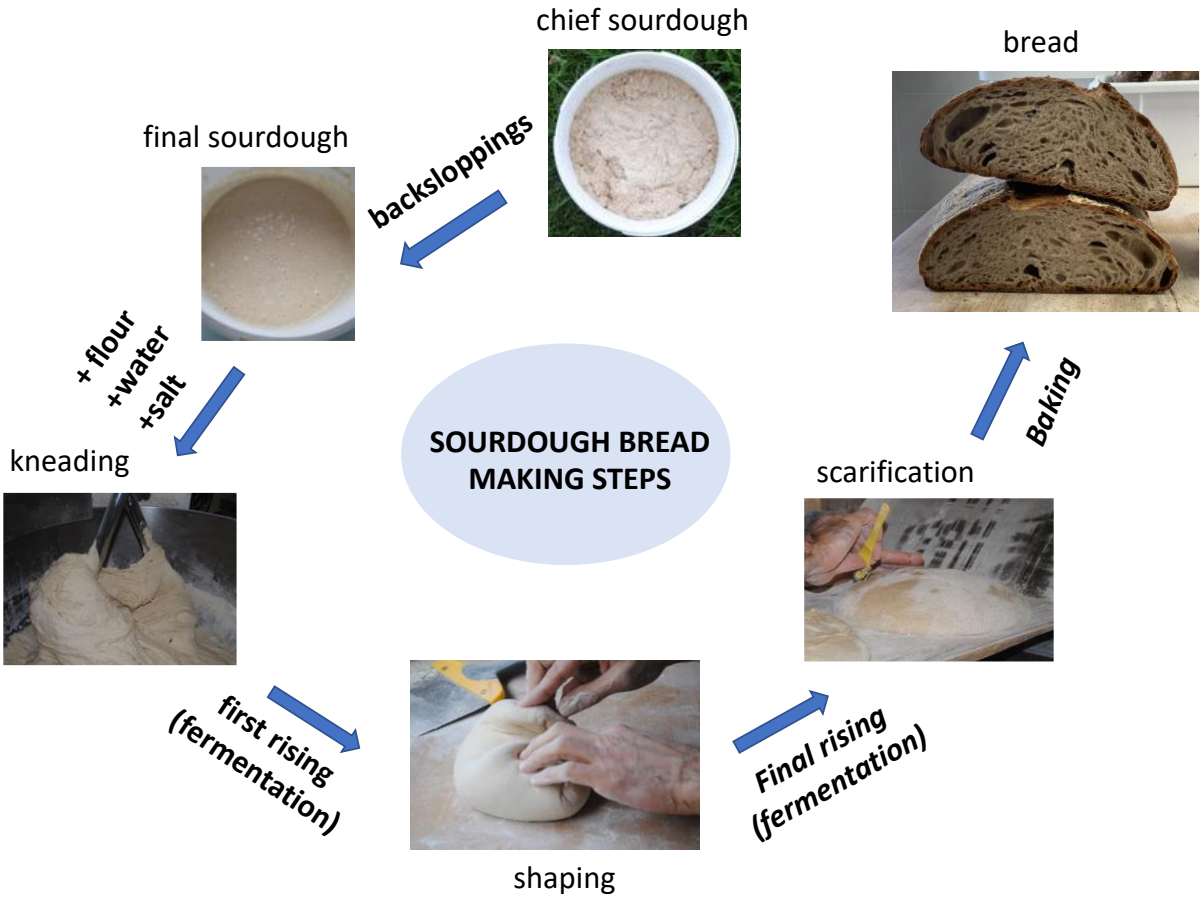

chief sourdough
bread
final sourdough
backsloppings
+ flour
+water
+salt
SOURDOUGH BREAD MAKING STEPS
Baking
kneading
scarification
first rising (fermentation)
Final rising (fermentation)
shaping

Figure 2

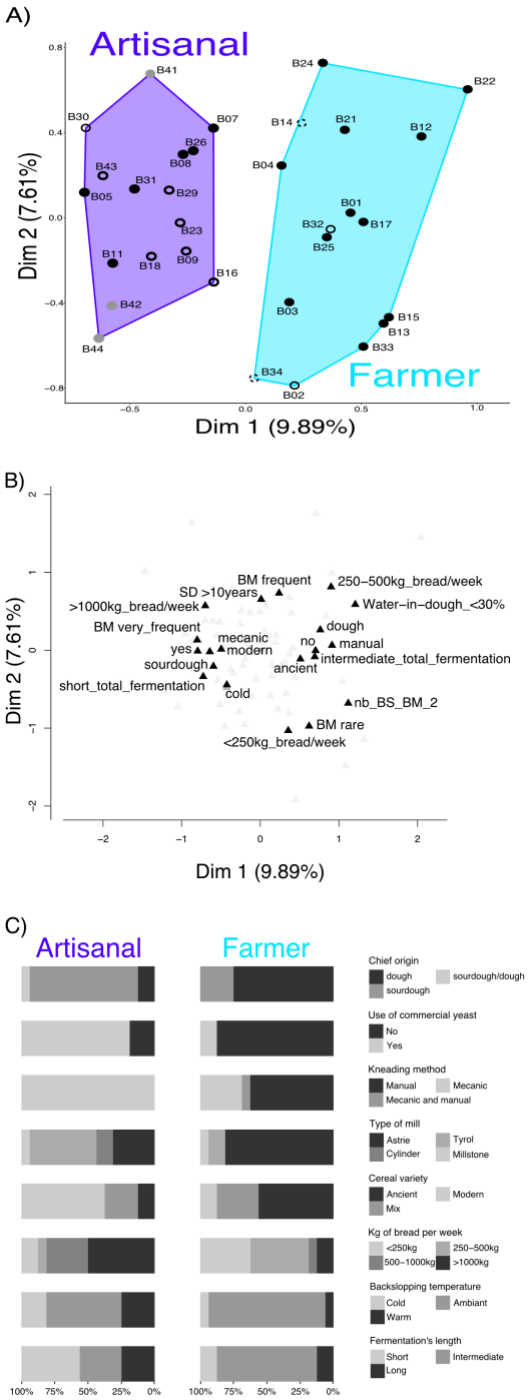

A)
Artisanal
Farmer
Dim 1 (9.89%)
Dim 2 (7.61%)
B)
BM frequent
250–500kg_bread/week
SD >10years
>1000kg_bread/week
Water–in–dough_<30%
BM very_frequent
dough
mecanic
yes
modern
no
manual
intermediate_total_fermentation
sourdough
ancient
short_total_fermentation
cold
nb_BS_BM_2
BM rare
<250kg_bread/week
Dim 1 (9.89%)
Dim 2 (7.61%)
C)
Artisanal
Farmer
Chief origin
dough
sourdough
sourdough/dough
Use of commercial yeast
No
Yes
Kneading method
Manual
Mecanic
Mecanic and manual
Type of mill
Astrie
Tyrol
Cylinder
Millstone
Cereal variety
Ancient
Modern
Mix
Kg of bread per week
<250kg
250–500kg
500–1000kg
>1000kg
Backslopping temperature
Cold
Ambiant
Warm
Fermentation's length
Short
Intermediate
Long
100%
75%
50%
25%
0%

Figure 3

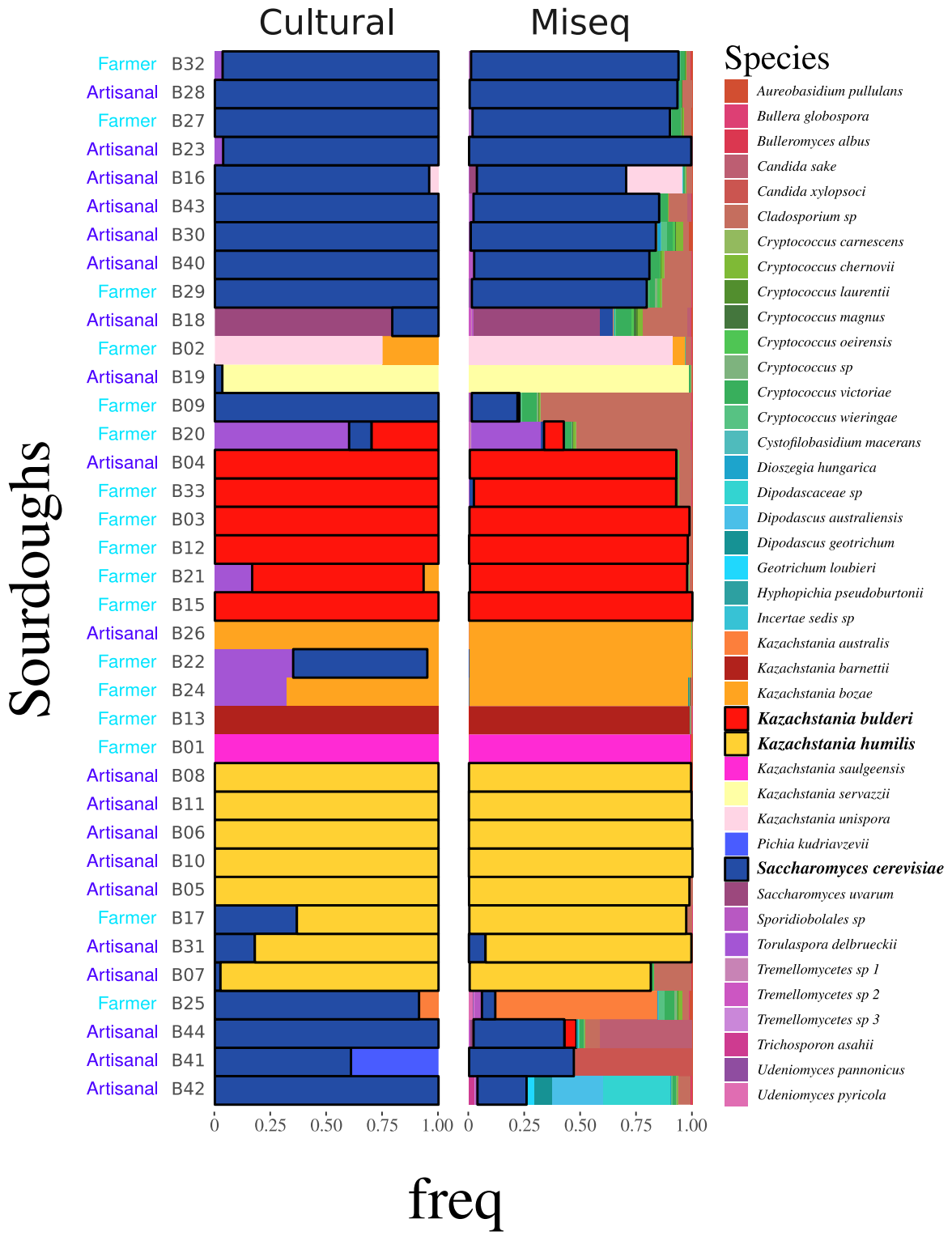

Cultural
Miseq
Sourdoughs
Farmer B32
Artisanal B28
Farmer B27
Artisanal B23
Artisanal B16
Artisanal B43
Artisanal B30
Artisanal B40
Farmer B29
Artisanal B18
Farmer B02
Artisanal B19
Farmer B09
Farmer B20
Artisanal B04
Farmer B33
Farmer B03
Farmer B12
Farmer B21
Farmer B15
Artisanal B26
Farmer B22
Farmer B24
Farmer B13
Farmer B01
Artisanal B08
Artisanal B11
Artisanal B06
Artisanal B10
Artisanal B05
Farmer B17
Artisanal B31
Artisanal B07
Farmer B25
Artisanal B44
Artisanal B41
Artisanal B42
0
0.25
0.50
0.75
1.00
0
0.25
0.50
0.75
1.00
freq
Species
Aureobasidium pullulans
Bullera globospora
Bulleromyces albus
Candida sake
Candida xylopsoci
Cladosporium sp
Cryptococcus carnescens
Cryptococcus chernovii
Cryptococcus laurentii
Cryptococcus magnus
Cryptococcus oeirensis
Cryptococcus sp
Cryptococcus victoriae
Cryptococcus wieringae
Cystofilobasidium macerans
Dioszegia hungarica
Dipodascaceae sp
Dipodascus australiensis
Dipodascus geotrichum
Geotrichum loubieri
Hyphopichia pseudoburtonii
Incertae sedis sp
Kazachstania australis
Kazachstania barnettii
Kazachstania bozae
Kazachstania bulderi
Kazachstania humilis
Kazachstania saulgeensis
Kazachstania servazzii
Kazachstania unispora
Pichia kudriavzevii
Saccharomyces cerevisiae
Saccharomyces uvarum
Sporidiobolales sp
Torulaspora delbrueckii
Tremellomycetes sp 1
Tremellomycetes sp 2
Tremellomycetes sp 3
Trichosporon asahii
Udeniomyces pannonicus
Udeniomyces pyricola

Figure 4

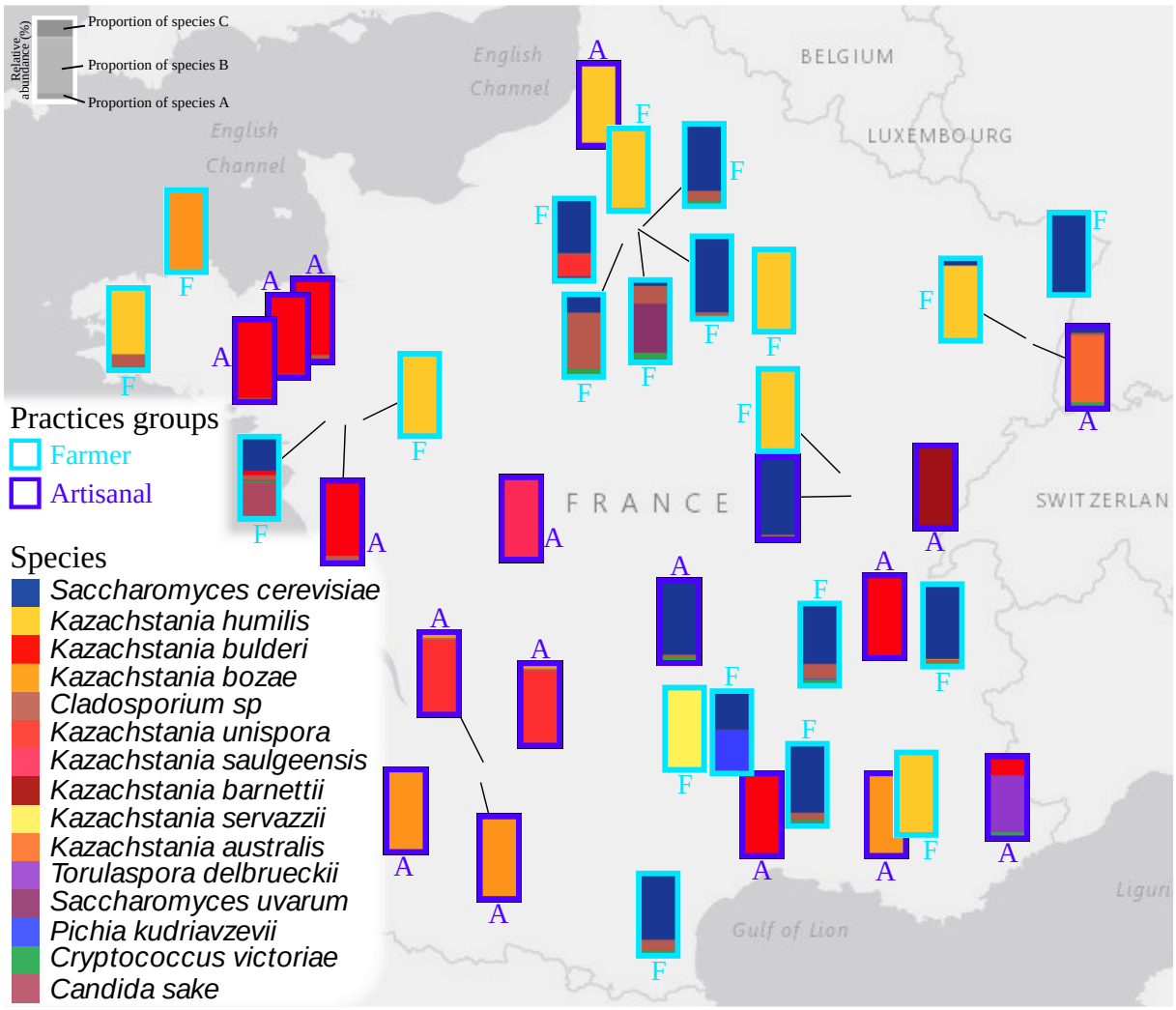
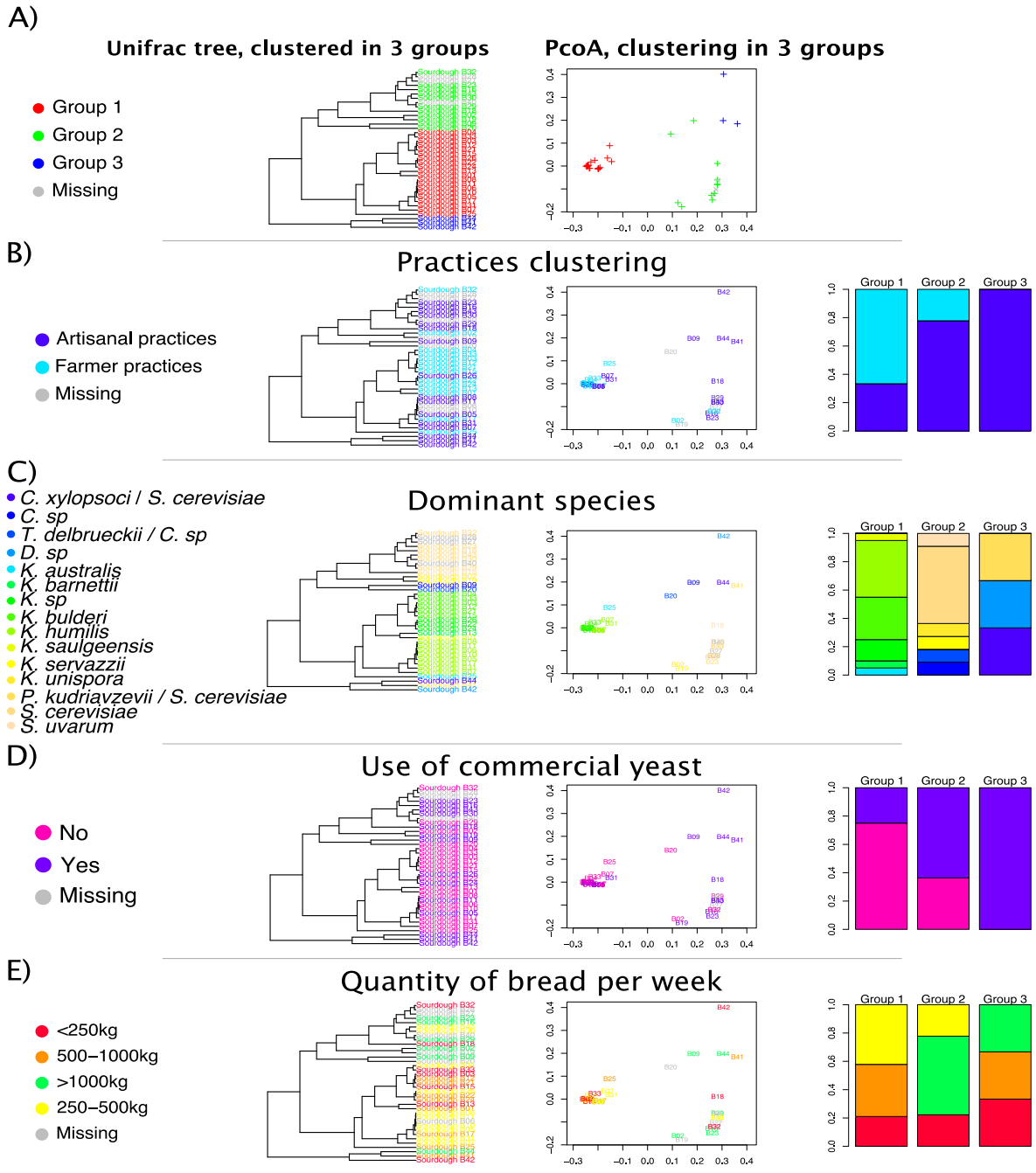
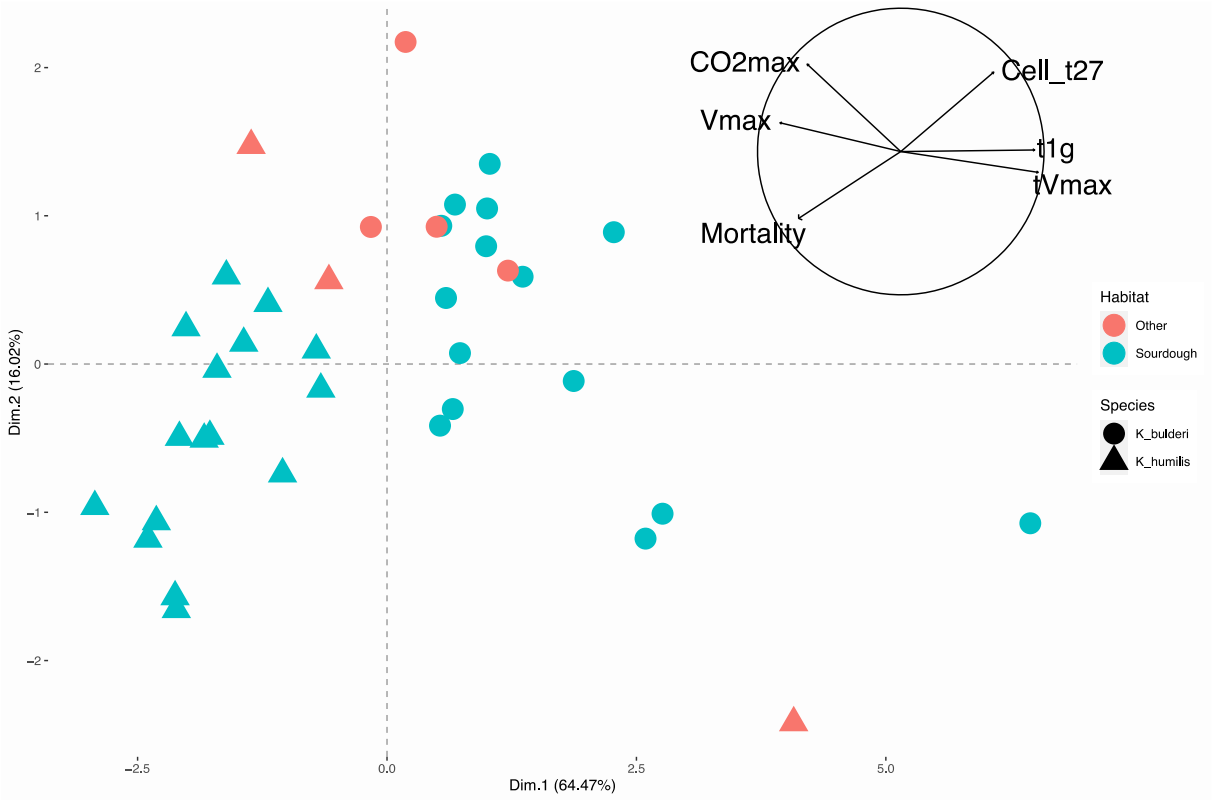

Relative abundance (%)
Proportion of species C
Proportion of species B
Proportion of species A
English Channel
English Channel
BELGIUM
LUXEMBOURG
FRANCE
SWITZERLAN
Gulf of Lion
Practices groups
Farmer
Artisanal
Species
Saccharomyces cerevisiae
Kazachstania humilis
Kazachstania bulderi
Kazachstania bozae
Cladosporium sp
Kazachstania unispora
Kazachstania saulgeensis
Kazachstania barnettii
Kazachstania servazzii
Kazachstania australis
Torulaspora delbrueckii
Saccharomyces uvarum
Pichia kudriavzevii
Cryptococcus victoriae
Candida sake

Figure 5


A)
Unifrac tree, clustered in 3 groups
PcoA, clustering in 3 groups
Group 1
Group 2
Group 3
Missing
B)
Practices clustering
Artisanal practices
Farmer practices
Missing
Group 1
Group 2
Group 3
C)
Dominant species
C. xylopsoci / S. cerevisiae
C. sp
T. delbrueckii / C. sp
D. sp
K. australis
K. barnettii
K. sp
K. bulderi
K. humilis
K. saulgeensis
K. servazzii
K. unispora
P. kudriavzevii / S. cerevisiae
S. cerevisiae
S. uvarum
D)
Use of commercial yeast
No
Yes
Missing
E)
Quantity of bread per week
<250kg
500–1000kg
>1000kg
250–500kg
Missing

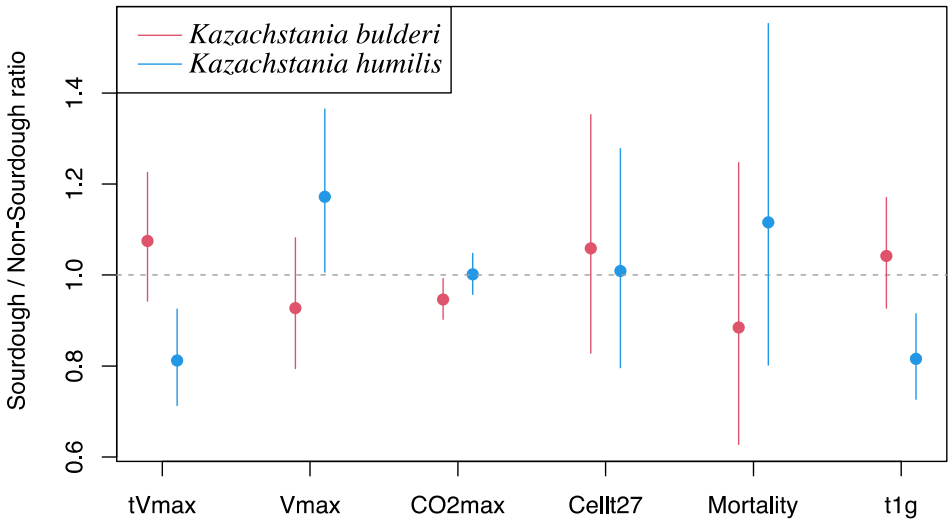
941

Figure 6


CO2max
Cell_t27
Vmax
t1g
tVmax
Mortality
Habitat
Other
Sourdough
Species
K_bulderi
K_humilis
Dim.2 (16.02%)
Dim.1 (64.47%)
2
1
0
−1
−2
−2.5
0.0
2.5
5.0

945 Figure 7

946



947