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[‡], 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

[¶]MaIAGE, INRA, Université Paris-Saclay, 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

29

30 No competing interests

31

32 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 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, psycho-sociologists and microbiologists to analyse French bread making practices and describe fungal communities in naturally fermented sourdough of 27 bakers and 12 farmer bakers. Their bread making practices were classified in two groups: the farmer-like practices group and the artisanal-like 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 while other yeast species of the closely related *Kazachstania* genus were dominant in 54% of sourdoughs. Bread making practices were found to drive the distribution of yeast species. The most remarkable difference 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. Phenotyping of these two species in laboratory sourdough mimicking media revealed signatures of domestication for *K. humilis* but not for *K. bulderi*. Overall, our results showed that preserving bread making practices diversity allows the preservation of a higher taxonomic and phenotypic 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 [1–3]. Domestication of the yeast *Saccharomyces cerevisiae* for the production of beer, wine, cheese, leavened bread, that of the fungi *Penicillium roquefortii* or *Penicillium camemberti* for cheese production or that of the fungus *Aspergillus oryzae* for rice or soybean fermented products are well studied cases [1–10]. The recent renewed interest in artisanal practices that make use of naturally fermenting microbial communities could promote again the conservation of microbial diversity. However, the effect of artisanal practices on the distribution of microbial species diversity remains poorly documented.

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 [11]. Since the Neolithic, bread history is intimately combined with the evolution of cereals, bread making associated tools and the advent of Mediterranean civilizations [12]. Leavened bread was traditionally made with flour, water and a fermenting agent, which was either a fermenting beverage or a fermenting dough, called sourdough. This sourdough was generally initiated from a mixture of flour and water, naturally colonized by lactic acid bacteria (LAB) and yeasts. Sourdough was then either maintained from one bread making process to the other or initiated again and again, depending on the craftsman [13, 14] . In the 19th century, the use of yeast starters made of *S. cerevisiae,* often called « baker's yeast », spread as an alternative to sourdough. Nowadays, *S. cerevisiae* industrial starters are more frequently used than sourdoughs, although the latter are gaining interest. A recent study showed that industrial populations of *S. cerevisiae* have followed a different evolutionary path than sourdough populations of *S. cerevisiae* [10]. Both have been

domesticated by humans and have improved their fermentation performance in a sourdough-mimicking medium. However, industrial and sourdough strains of *S. cerevisiae* differ genetically and phenotypically, indicating that sourdough bread making contributes to the conservation of *S. cerevisiae* diversity [10].

Beyond *S. cerevisiae*, sourdough can also host other yeast species. 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 that may determine the distribution 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, grain intake, house characteristics [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 a professional baker. To our knowledge, no studies have been conducted to date to investigate the effect of bakers' bread-making practices on sourdough yeast diversity.

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 [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 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.

105 # Materials and methods

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

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

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

120 **Sourdoughs samples, enumeration and strain isolation**

121 Sourdoughs were collected before kneading and referred to as final sourdoughs (Table S1). On the day
122 of collection, they were sent to the lab where yeast and bacteria were enumerated and isolated as in [19,
123 20], and sourdoughs stored at -20°C in sterile vials for non- culture based analysis. Ethics and rights
124 associated with sourdough collection and strains isolation have been respected.

125 **Sourdough acidity and metabolic analyses**

126 For each sourdough, three independent 1-g replicates were analysed. pH and Total Titrable Acidity were
127 measured as described in [20]. Organic acids, alcohol and sugars concentrations (expressed as g/kg of
128 sourdough) were analysed by liquid chromatography using an HPLC HP 1100 LC system (Agilent

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

134 **Yeast species identification**

135 The Internal transcribed spacer 1 (ITS1) ribosomal DNA of each 1216 yeast isolates was amplified by
136 PCR from chromosomal DNA, either by using primers ITS1F and ITS2 [21, 22], or primers NSA3 and
137 58A2R [19, 22]. For isolates unidentified with the ITS1 region alone, DNA was extracted according to
138 MasterPure Yeast DNA purification kit (Epicentre, Epibio). PCR reactions targeting partial genes, the
139 D1D2 region of the large subunit of rRNA (LSU), a part of the RNA polymerase II large subunit
140 encoding gene (*RPB1)*, a part of the RNA polymerase II encoding gene (*RPB2)*, a part of the actin
141 encoding gene *Act1* and Transcriptional Elongation Factor TEF were performed. To discriminate three
142 specific isolates, PCR on genes *GHD1*, *FSY1*, *URA3*, *DRC1*, *MET2* were performed [23–26] (Table S2).
143 All PCR products were sent to be sequenced with Sanger sequencing (Eurofins, Germany). Species were
144 identified using NCBI [27], YeastIP [28] and a personal database, which was constructed after *ITS1*,
145 *RPB2*, LSU sequencing of all 33-yeast species reportedly found in sourdoughs in the literature [19].

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

147 The ITS1 region was targeted with the PCR primers ITS1-F (5′- CTTGGTCATTTAGAGGAAGTAA -
148 3′) and ITS2 (5’-GCTGCGTTCTTCATCGATGC-3’).

149 The sequencing run was performed with MiSeq Reagent Kit v3. 2015 [20]. Sequences were analysed
150 through FROGS “Find Rapidly OTU with Galaxy Solution” [29] and home-made pipelines. Overlapped
151 reads were merged with Flash [30] with a minimum overlap of 10 nucleotides, a maximum overlap of
152 300 nucleotides and a maximum mismatch density of 0.1. Adapters were removed with Cutadapt [31]
153 and data were cleaned with Sickle [32]. Reads were clustered with Swarm [33] and chimeras deleted

154 with VSEARCH [34]. Sequences were then filtered on minimum abundance of 0.05% of all sequences.
155 From the OTU's abundance table and for each OTU, the taxonomic affiliation using UNITE Version
156 7.1, Release 2016-11-20 [35], YeastIP [28] and our own databases [19] was obtained by blasting OTUs
157 representative sequences against each database.

158 **Phenotypic analysis of yeast strains**

159 Fermentations performance of the two most frequently encountered Kazachstania species were assessed
160 as described in [10] for *S. cerevisiae*. 15 sourdough strains of *K. bulderi* and 16 sourdough strains of *K.*
161 *humilis* were included in the analysis. *K. bulderi* strains were coming from sourdoughs B4, B12, B15,
162 B17 B20, B21, while *K. humilis* strains were coming from sourdoughs B2, B5, B6, B7, B10, B17. From
163 one to three strains per sourdough were analysed in the experiment. In addition, four strains of *K. bulderi*
164 (strain MUCL 38021 isolated from ensilage in Namur, Belgium, strain MUCL 54694 isolated from
165 ensilage in Erezée, Belgium, strain NRRL Y-27205 isolated from three month old maize silage in the
166 Netherland, and strain CLIB 604 coming from Agro food industry) and three strains of *K. humilis* (strain
167 CBS 7754 isolated from food dressing, strain CBS1323 isolated from bantu beer, and strain CBS 2664
168 isolated from alpechin) which were coming from non-sourdough habitats, were added as control to test
169 the effect of habitat of origin. Each strain was phenotyped at least in triplicate leading to a total of 145
170 fermentations distributed over two blocks. Briefly, fermentations were carried out at 24 °C with constant
171 magnetic stirring (300 rpm) during 24 h. CO_2 release was measured by weight loss every 40 min using
172 an automated robotic system [36]. At the end of fermentation, population size and cell viability were
173 determined by flow cytometer (C6 cytometer, Accuri, BD Biosciences) as described in [37].

174

175 **Data analyses**

176 To analyse bread making practices, a multiple correspondence analysis (MCA) and hierarchical
177 clustering (complete linkage clustering method) on principal components based on the first two axes of
178 the MCA were performed using the FactoMineR R package [38].

179 To analyse fungal community, weighted Unifrac distances between sourdough communities were
180 computed from a rooted phylogenetic tree based on the OTUs sequences using the R-packages Phyloseq
181 and GUniFrac [39, 40]. Phylogenetic sequences were aligned with Clustalo and phylogenetic tree were
182 built with the parsimony algorithm, with 100 replicates bootstraps, pairwise ktuple-distances with
183 Seaview [41]. Different roots were tested (*Sporidiobolales sp.*, *Bullera globospora*, *Trichosporon*
184 *asahii*, *Udeniomyces pyricola*). Tree architecture did not change with the root. It did not fit the expected
185 phylogeny and, notably, some *Ascomycota* were misclassified among the *Basidiomycota*. However, the
186 dominant sourdough species belonging to the *Saccharomycetaceae* family were clustered according to
187 expected clades or subclades, except that *Kazachstania servazzi* and *Kazachstania unispora* were
188 grouped in a clade closer to *Saccharomyces* species than to other *Kazachstania* species. Using the
189 Unifrac distances matrix, we performed a Principal Coordinate Analysis (PCoA) and clustered
190 sourdough communities using the first two axes of the PCoA, and the complete linkage clustering
191 method (hclust R function). To check the sensitivity of our analysis to this misclassification, we
192 performed the same analyses without the sourdoughs that had one misclassified species representing
193 more than 10% of their reads, *i.e.* sourdoughs B20, B41, B42, and B44 and found the same clustering
194 [40].

195 For each sourdough, the species richness, Chao1, Shannon and Simpson indexes were computed, and
196 the Shannon and Simpson indices values were converted to the effective number of species per
197 sourdough. This number was estimated from the Shannon diversity index as $exp^{Shannon\ index}$ and from
198 the Simpson diversity index as $\frac{1}{1-Simpson\ index}$ [42, 43]. For probability estimates, the exact 95%
199 confidence intervals were computed using a binomial distribution.

200 To study the link between microbial diversity and bakery practices, a univariate Permutational
201 Multivariate Analysis (PERMANOVA) on the Unifrac distance matrix was performed for each bakery
202 practice variable. We performed univariate analysis on the 30 bakers who had less than 8 missing values
203 among the 29 bread making practices variables and adjusted the p-value using FDR correction to account
204 for multiple testing. In addition, independence exact Fisher tests between the variable providing fungal

205 community PCoA groups and each of the bread making practices variables were performed. Multiple
206 testing was accounted for using the False Discovery Rate method [44].

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

219 Finally, the phenotypic diversity of *K. bulderi* and *K. humilis* strains coming from sourdough and non-
220 sourdough habitat were analyzed. Population size and mortality rate after 27h of fermentation measured
221 by flow cytometer were used as proxy for absolute fitness. The cumulative CO_2 production curve was
222 calculated and the kinetics of CO_2 production rate over time was estimated by successive linear
223 smoothing over 5 points. Four fermentation parameters were then estimated: the maximum CO_2 release
224 (CO_2max, in gr/L) was estimated by the maximum of the cumulative CO_2 production curve, the
225 fermentation latency-phase time (t1g, in h) was estimated by the time between inoculation and the
226 beginning of the fermentation calculated as 1g/L of CO_2 release, the maximum CO_2 production rate
227 (Vmax in g/L/h) was estimated by the maximum of the CO_2 production rate kinetic and the time of the
228 Vmax (tVmax in h) as the time between inoculation and the Vmax. Hence, the phenotype of each strain
229 was characterized by six quantitative variables called "phenotype variables" below: its population size,
230 its mortality rate and the four fermentation parameters. To determine whether the origin of the strain

(sourdough or non-sourdough) had an impact on strain phenotype, each log-transformed quantitative variable was analyzed separately using a mixed linear model as described below. The experimental design was unbalanced between the two blocks with very few non-sourdough strains in one of the two blocks. Therefore, for each phenotype variable, we first estimated the block effect with a subset of 8 strains cultivated in both blocks using a linear model with two fixed effects : the strain and the block. Additive models were used as the interactions terms were not significantly different from zero after adjusting p-values with the Benjamini-Hochberg method. Second, each phenotype variable was corrected for the block effect and analyzed with the mixed effect model: $Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_{ij} + Z_k + \epsilon_{ijkl}$ where Y_{ijkl} represents the log-transformed phenotype variable corrected for the block effect for the strain *k*, from species *i (i=1,2)*, sampled in environment *j (j=1,2)*, observed for replicate *l*. μ represents the mean of the phenotype variable, α_i the additive effect of species i, β_j the additive effect of environment *j*, and γ_{ij} their interaction. Z_k represents the gaussian random effect of strain *k* with $Z_k \sim N(0, \sigma_s^2)$ and ϵ_{ijkl} the gaussian residuals with $\epsilon_{ijkl} \sim N(0, \sigma^2)$. For each species i, the impact of the environment was quantified using the contrast $\Delta_i = \beta_S + \gamma_{i,S} - \beta_{NS} - \gamma_{i,NS}$ with NS standing for "non-sourdough" and "S" for '"sourdough" and tests ($H_0 = \Delta_i = 0, H_1: \Delta_i \neq 0$) were performed and 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 (ggplot2 [48], leaflet package [49], with minor esthetical adjustment with Inkscape) have been done with R. Data and scripts are shared on Zenodo. http://doi.org/10.5281/zenodo.2600170

253 # Results and discussion

254 ## Bread making processes and the diversity of bread making practices

255 A total of 39 French bakers producing natural sourdough bread and distributed all over France
256 participated to the study (Table S1). The bread making practices of 35 of them were collected through
257 one or several methods: personal interviews (with 12 bakers), focus groups (3 groups), observation
258 during break-making workshops (2 workshops), and an online/phone survey (36 bakers). The general
259 process of sourdough bread making is presented in Figure 1. We analysed 28 variables around this
260 general bread making process, describing variations of the practices at all steps of the bread making
261 process, from wheat grains to baked bread (Figure S1). Four bakers (B6, B10, B19, B20) who did not
262 provide enough information about their practices were excluded from the multivariate analysis.
263 According to a hierarchical clustering on principal components (HCPC), the 32 other bakers clustered
264 into two groups corresponding to two main types of bread making practices (Figure 2). The first group,
265 hereafter termed "farmer-like" practices group, includes 6 bakers and 11 farmers-bakers using the
266 following practices: low bread production (<500 kg per week, 81% of the bakers of the "farmer-like"
267 group), use of ancient wheat populations (56%), manual kneading (63%), working at ambient
268 temperature (88%), long fermentation periods (more than 4 hours for 88%), and no use of commercial
269 baker's yeast (88%). In addition, they tend to make their chief sourdough from dough after kneading
270 (75%). The second group, hereafter called "artisanal-like" practices group consists of 12 bakers and 4
271 farmer-bakers having more intensive practices, characterized by a large bread production (>500 kg per
272 week, 81%), mechanical kneading (100%), use of modern wheat varieties (63%), working at ambient
273 temperature (56%), using commercial yeast starters in addition to sourdough for bread making or using
274 commercial yeast starters for pastries and buns making (81%). In this second group, bakers tend to make
275 their chief sourdough from a final sourdough.

276 **Fungal community composition**

277 Sourdough is a mix of flour and water naturally fermented by bacteria and yeasts. Sourdough yeast
278 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
279 per gram, as commonly found in sourdoughs from all over the world [1, 16, 50, 51]. We isolated 20 to
280 40 yeast strains from each sourdough and identified species using ITS sequence as well as other barcodes
281 when needed (see M&M section). Among the 39 collected sourdoughs, one (sourdough B14) did not
282 give any colony in the laboratory, suggesting that his sourdough microbiota was no longer alive. A total
283 of 1216 strains were characterized from the other 38 sourdoughs. In addition, we developed an ITS1
284 meta-barcoding MiSeq sequencing method on sourdough (see sup M&M). After filtering 5,360,620 raw
285 ITS1 sequences for quality, abundance (0.005%) and chimera, 3,542,801 sequences were further
286 analyzed. Overall, the sequences clustered in 113 OTUs. The number of reads per sourdough ranged
287 from 8421 to 194 557. Therefore, we carried out our analysis on the rarefied matrix. Among all OTUs,
288 64 were identified as non-yeast, including 10 assigned to the order *Triticodae* (especially to the species
289 *T. aestivum*), 50 assigned to plant pathogen fungi, such as *Alternaria*, *Aspergillus* or *Fusarium*,
290 *Gibberella*, while 4 OTUs remained unidentified. Among the 40 yeast or yeast-like OTUs, 96% of total
291 reads were assigned to the phylum *Ascomycota*, 87.5% to the order *Saccharomycetales* and 85.7% to
292 the family *Saccharomycetaceae*. Only 4% of the total reads were assigned to the phylum *Basidiomycota*.
293 Overall, three OTUs assigned to the species *Kazachstania humilis*, *Kazachstania bulderi* and
294 *Saccharomyces cerevisiae* represented 20.3%, 15.5% and 24.1% respectively of the total number of
295 reads and 28.1%, 23.7% and 18.2% respectively of the number of reads identified as yeast species
296 (Figure 3).

297 Both non-culture-based and culture-based methods allowed the identification of the same dominant
298 species (defined as a species with an over 50% frequency) for all sourdoughs but five (B09, B20, B22,
299 B25, B41) (Figure 3). In two cases, the discrepancy was explained by the detection of *Cladosporium* sp.
300 at high frequency with metabarcoding while this species could not be isolated in the laboratory (Figure
301 3). In two other cases, it was explained by a high number of *S. cerevisiae* isolated in the laboratory

302 compared to what was observed using metabarcoding sequencing. Finally, in the last case, the
303 identification of *Pichia kudriavzevii* required additional sequencing as it shares an identical ITS with
304 *Candida xylopsoci*. Because metabarcoding allows a deeper characterization of the fungal species
305 diversity with few discrepancy cases, the distribution of fungal species diversity will be further described
306 using metabarcoding data only. Previous analysis of the same sourdoughs revealed that *F.*
307 *sanfranciscensis* was the dominant bacterial species in all analysed sourdoughs but two, where the
308 dominant species was either *L. curvatus* or *L. heilongjiangensis* [20, 52, 53]. Therefore, we decided to
309 study the distribution pattern of microbial species in the fungal community only.

310 **Distribution of fungal species diversity over sourdoughs**

311 All sourdoughs but two had a dominant yeast species with a relative abundance over 50% and
312 many species with a lower relative abundance (Figure 3). Within sourdoughs, fungal species richness
313 ranged from 10 to 33, with a 23.5 median (Table S3). The effective number of species per sourdough
314 calculated from the Shannon diversity index ranged from 1 to 6.8 (Table S3), with 70% of sourdoughs
315 having an index below 2 (Table S3). Between-sourdoughs species diversity was analysed using weighted
316 Unifrac distances, computed from a phylogenetic tree built from the distances between OTUs using
317 *Sporidiobolales species* as root (Figure S2). Unifrac distances computed with four differently rooted
318 trees were highly positively correlated (Figure S3). Unifrac distances between sourdoughs ranged from
319 0.0005 and 0.71, with a median of 0.49 and a mean of 0.52. There was no significant correlation between
320 the Unifrac distances and geographical distances between sourdoughs (Mantel test, P=0.35).

321 We then analysed specifically the distribution of yeast species diversity as yeast species,
322 together with lactic acid bacteria, are the main functional player in a sourdough ecosystem and for bread
323 quality. Over the 40 yeast species detected in the 38 sourdoughs, 12 had a relative abundance over 50%
324 in at least one sourdough, four had a relative abundance between 20% and 50% and 24 had a relative
325 abundance below 10%. All dominant species (relative abundance over 50%) were fermentative yeast
326 species, except in one sourdough that had a C*ladosporium* species. We found all the sourdough yeast

327 genera (*Saccharomyces*, *Kazachstania*, *Pichia*, *Torulaspora* and *Hyphopichia)* commonly reported in
328 the literature except the *Wickerhamomyces* genus that we did not detect in our samples [1, 15, 16].

329 The baker's yeast species, *Saccharomyces cerevisiae* is not the most widespread yeast 330 species in French organic sourdoughs

331 *S. cerevisiae* was found in 53% of all sourdoughs (95% confidence intervals=36% - 69%) but was
332 dominant (relative abundance over 50%) in only 24% (95% confidence intervals=11% - 40%) (Figure
333 3). In two cases, *S. cerevisiae* co-occurred with another yeast species at similar relative abundance. In
334 the first case, *S. cerevisiae* was present at a relative abundance of 40% with *Candida sake* at a 41%
335 relative abundance. In the second case, it was found at a relative abundance of 47% with *Pichia*
336 *kudriavzevii* at a relative abundance of 52%. In all the other cases, *S. cerevisiae* had a relative abundance
337 below 21% and was found with other dominant yeast species, such as *Kazachstania australis*,
338 *Kazachstania humilis*, *Saccharomyces uvarum* or *Torulaspora delbrueckii*. This suggests that *S.*
339 *cerevisiae* did not displace other species and can indeed be out-competed by other species in sourdoughs.
340 Therefore, despite the recurrent use of *S. cerevisiae* as starter species, its massive use in some bakeries
341 and homes, and its occurrence in a wide range of habitats such as soil, trees, and humans [9, 54] , this
342 species does not appear to have overwhelmingly colonized French traditional sourdoughs (Figure 3,
343 Figure 4). This result confirmed a recent analysis which revealed that *S. cerevisiae* sourdough strains
344 have a different evolutionary history than industrial strains [10].

345 The *Kazachstania* genus is highly represented in sourdoughs

346 *Kazachstania* was the most represented yeast genus over all sourdoughs, when considering both the
347 number of reads over all sourdoughs and the number of detected species. Indeed, this genus represented
348 57% of the total number of reads while *Saccharomyces* represented 26% of the total number of reads.
349 In addition, eight species of the *Kazachstania* genus were found in sourdough, a much higher figure
350 than for the *Saccharomyces* genus, represented by only two species (*S. uvarum* and *S. cerevisiae*) (Figure
351 3). *Kazachstania* species dominated in 54% (95% confidence intervals=36%-69%) of sourdoughs while

352 *Saccharomyces* species dominated in 27% only (95% confidence intervals=13%-43%). *K. humilis*,
353 followed by *K. bulderi* were the most commonly dominant *Kazachstania* species, and found in
354 respectively 21% (95% confidence intervals=10%-37%) and 15% of sourdoughs (95% confidence
355 intervals=6%-31%) (Figure 3). *K. humilis* is common in sourdoughs and has been found in many
356 countries, viz. Austria, Canada, China, Denmark, Ethiopia, Finland, Germany, Greece, Italy, Morocco,
357 the Netherland, Spain, UK, USA, and France [1, 15–17, 19, 53, 55–60]. *K. bulderi* has been reported
358 for the first time in anaerobic maize silage in the Netherlands and in fermented liquid feed for piglets
359 [61, 62] and more recently in French, Belgium and Spain sourdoughs [19, 53, 57, 63]. A recently
360 described *Kazachstania* species, *Kazachstania bozae*, was also identified in five sourdoughs (4.5%-
361 29%) and found dominant in three (1.7%-22%) [64]. Strains of this species were closely related to a
362 strain previously isolated from boza, a Bulgarian fermented drink, as estimated with ITS and LSU
363 (D1D2) barcodes (Source: NCBI, GenBank: KC118125.1 and KX369579.1). In addition, *Kazachstania*
364 *saulgeensis,* a recently described species [65, 66], was dominant in one sourdough (0.07%-14%).
365 *Kazachstania unispora* and *Kazachstania servazzi* which had previously been detected in sourdough
366 were also found [17, 17, 53, 57, 58, 63, 67–71]. Finally, some *Kazachstania* species were detected for
367 the first time as dominant in sourdoughs, whereas they had been previously found in other environments,
368 like soil (*K. australis*), sauerkraut (*K. barnettii*) [72–74]. None of the previous studies on sourdough
369 have evidenced as many *Kazachstania* species in sourdough.

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

372 We tested whether sourdough fungal community composition could be explained by bread making
373 practices. To do so, we performed univariate PERMANOVA analysis on the 30 bakers with less than 8
374 missing values for the 29 bread making practices variables (Table S4). The univariate analysis revealed
375 that the weighted Unifrac distance is structured according to the use of commercial yeast in bakery
376 ($P<0.05$). It also varied significantly with sourdough age, chief sourdough origin (dough, sourdough or
377 both), the quantity of bread produced per week, the milling method (cylinder, millstone, Astrie, Tyrol),
378 the type of wheat variety (ancient, modern or a mix thereof) and the fermentation duration. However,

379 after FDR correction for taking into account multiple testing, none of these variables significantly
380 explained Unifrac distances.

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

403 To test more specifically the link between bread making practices and the distribution of
404 *Kazachstania* species, we analyzed more in depth group 1 sourdoughs. Within this group, 8 sourdoughs
405 had *K. humilis* as dominant species, 6 had *K. bulderi*, 3 had *K. bozae* and the remainder had still other

406 *Kazachstania* species. All sourdoughs made with artisanal practices carried *K. humilis* as dominant
407 species or, in one case, the *K. bozae*. By contrast, sourdoughs made with farmers' practices had as
408 dominant species *Kazachstania bulderi*, *K. australis*, *K. barnettii*, *K. saulgeensis* or *K. bozae* (exact
409 Fisher test, P=0.004).

410 Other bread making practices than the ones studied here could also explain the distribution of yeast
411 species diversity. Interviews with the bakers working with sourdough hosting *K. bulderi* and *K. bozae*
412 suggested the role of dispersion of these species in French sourdoughs. Indeed, these bakers have been
413 connected over the years either through seed exchanges, sourdoughs mixing or gifts, bread making
414 training in common or working in one another's bakery. Some yeast species have been found in the
415 bakery house environment and baker's hands and may thus be dispersed through baker's tools or baker's
416 travels [58, 67, 75, 76]. However, it is still unclear whether wheat seeds and flour are a source of
417 sourdough yeasts.

418

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

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

427 To investigate the relation between sourdough fungal communities and metabolic compounds, we
428 quantified sourdough hydration, yeast density, bacteria density, sourdough pH, Total Titrable Acidity
429 (TTA), sourdough concentration in seven sugars (maltose, glucose, fructose, raffinose, arabinose,
430 mannose, xylose), four alcohols (glycerol, ethanol, mannitol, meso-erythtritol), six acids (lactate,

431 acetate, glutarate, pyruvate, malate, succinate) and calculated the fermentative quotient (lactate over
432 acetate ratio). For each variable, there was a wide range of variation (Table S5). The principal component
433 analysis based on all variables showed no evidence of sourdoughs grouping (Figure S4). As expected
434 in fermentation, yeast density was positively correlated to ethanol (r=0.74, P<0.001), glycerol (r=0.67,
435 P<0.001), and acetate (r=0.6, P<0.001) concentration. However, it was not significantly correlated to
436 sugar concentrations. This might be explained by the co-occurrence of bacteria which have their own
437 metabolism and interact by competition and/or cross feeding with sourdough yeasts.

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

444 In addition, we tested whether variations in quantitative variables were associated with the fungal
445 community groups (Table S5). Group 3 microbial community sourdoughs (defined by PCoA clustering
446 on Unifrac distance, see below), which contains *S. cerevisiae* in co-dominance with a second yeast
447 species (*Candida sake*, *Pichia kudriavzevii* or a *Dipodascus* species), had a significantly higher mean
448 pH (mean pH_{group3}=4.2 against pH_{group1}=3.8, Tukey Contrasts, P<0.001), lower TTA (mean TTA_{group3}=7.7
449 against TTA_{group1}=17.1, Tukey Contrasts, P=0.002), and a higher maltose concentration (mean Maltose
450 $_{group3}$=52.8 mg/gr of sourdough against Maltose $_{group1}$=24.1 mg/gr of sourdough, Tukey Contrasts
451 P=0.002) than group1, having a *Kazachstania* dominant species. Compared to group 2 having in most
452 cases *S. cerevisiae* as dominant species, it also had higher pH (pH_{group2}=3.9, Tukey Contrasts, P=0.003),
453 and higher maltose concentration (Maltose $_{group2}$=23.7, Tukey contrast, P=0.003). These data may reflect
454 a lower fermentative activity for group 3 fungal community having two co-dominant species, and/or a
455 negative interaction effect of group3 fungal community on the activity of lactic acid bacteria (LAB),
456 which are the main producers of sourdough acids. Previous studies on the bacteria content of the same

457 sourdoughs showed that *L. sanfranciscensis* was most generally the dominant species, although *L.*
458 *heilongjiangensis, L. curvatus* or *L. brevis* were also found as dominant species [20, 52, 53]. We found
459 no significant correlation between LAB and yeast densities (r= -0.15, p= 0.45, Figure S4) but the link
460 between fungal and bacteria community might be species and strains dependent. Additional studies on
461 the interactions between fungal and bacterial communities need to be performed to better understand
462 how they may drive sourdough acidity and sugar content.

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

475 **Phenotypic signatures of domestication**

476 A previous analysis on *S. cerevisiae* revealed that sourdough strains had higher average fitness and
477 fermentation performance than strains from other environments in a sourdough-mimicking medium
478 (Bigey et al. 2021). Here, we investigated whether signatures of domestication could also be found in
479 *K. humilis* and *K. bulderi*, the two Kazachstania species most commonly found in French sourdoughs.
480 We tested whether fitness (log of population size and mortality at the end of fermentation) and
481 fermentation performance (CO2max, Vmax, t1g, tVmax) differed between sourdough strains and strains
482 from elsewhere.

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

491 Statistical comparison of sourdough and non-sourdough strains of *K. bulderi* and *K. humilis* for each
492 phenotypic variable revealed signatures of domestication for *K. humilis* but not for *K. bulderi*. While
493 the *K. bulderi* sourdough strains did not ferment significantly faster than the non-sourdough strains, the
494 *K. humilis* sourdough strains showed significantly higher Vmax, lower t1g, and lower tVmax than the
495 non-sourdough *K. humilis* strains (Figure 7, Table S6). On average, they started fermentation two hours
496 before the others and reached Vmax three hours before the others. In addition, their Vmax were on
497 average 34% higher than the others. *K. humilis* was found in French bakeries practicing artisanal bread
498 making. It is also the most frequently encountered Kazachstania species in sourdoughs around the world.
499 This species is therefore frequently found in bakeries, where short fermentations are often favoured.
500 This may explain why sourdough strains of *K. humilis* seem to have been selected for increased
501 fermentation speed. In contrast, sourdough-derived strains of *K. bulderi* did not appear to have a faster
502 fermentation rate than strains from other environments. *K. bulderi* was found in bakeries with farmer-
503 like practices. These bakeries often bake bread once or twice a week and store their sourdough for
504 several days. They also often use long fermentation and thus may not have selected an increased
505 fermentation rate. Other than fermentation phenotypes, there was no evidence of fitness differences
506 between sourdough and non-sourdough strains of *K. humilis* and *K. bulderi* in sourdough mimicking
507 media. Additional experiments in real dough should be performed to further test the effect of natural
508 selection in this environment.

Therefore, our results suggest that populations of *K. humilis* have been selected by humans and can therefore be considered domesticated. Selection by artisanal practices would have led to an increase in fermentation rate, which is a desired trait by humans to accelerate the bread-making process. In contrast, we did not find a domestication signature for *K. bulderi* found in sourdoughs made by farmer-bakers. Farmer-bakers typically store their sourdough for several days during the week and therefore make a lower number of backslopping during the week. 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.

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* 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. Bread-making practices divergence also led to different phenotypic signatures. Strains of the *K. humilis* species, which was typical of sourdough made by artisanal-like practices, had increased fermentation rate while strains of *K. bulderi* which was typical of sourdough made by farmer-like bread-making practices had not. To our knowledge, this is the first evidence of the influence of artisanal practices on taxonomic and phenotypic 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

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

539 **References**

540 1. Carbonetto B, Ramsayer J, Nidelet T, Legrand J, Sicard D. Bakery yeasts, a new model for
541 studies in ecology and evolution. *Yeast* ; **0**.
542 2. Gibbons JG, Rinker DC. The genomics of microbial domestication in the fermented food
543 environment. *Current Opinion in Genetics & Development* 2015; **35**: 1–8.
544 3. Steensels J, Gallone B, Voordeckers K, Verstrepen KJ. Domestication of Industrial Microbes.
545 *Current Biology* 2019; **29**: R381–R393.
546 4. Gibbons JG, Salichos L, Slot JC, Rinker DC, McGary KL, King JG, et al. The evolutionary
547 imprint of domestication on genome variation and function of the filamentous fungus Aspergillus
548 oryzae. *Curr Biol* 2012; **22**: 1403–1409.
549 5. Watarai N, Yamamoto N, Sawada K, Yamada T. Evolution of Aspergillus oryzae before and
550 after domestication inferred by large-scale comparative genomic analysis. *DNA Research* 2019; **26**:
551 465–472.
552 6. Ropars J, Didiot E, Rodríguez de la Vega RC, Bennetot B, Coton M, Poirier E, et al.
553 Domestication of the Emblematic White Cheese-Making Fungus Penicillium camemberti and Its
554 Diversification into Two Varieties. *Curr Biol* 2020; **30**: 4441-4453.e4.
555 7. Ropars J, López-Villavicencio M, Snirc A, Lacoste S, Giraud T. Blue cheese-making has
556 shaped the population genetic structure of the mould Penicillium roqueforti. *PLOS ONE* 2017; **12**:
557 e0171387.
558 8. Legras J-L, Galeote V, Bigey F, Camarasa C, Marsit S, Nidelet T, et al. Adaptation of S.
559 cerevisiae to Fermented Food Environments Reveals Remarkable Genome Plasticity and the
560 Footprints of Domestication. *Mol Biol Evol* 2018; **35**: 1712–1727.

561 9. Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, et al. Genome evolution
562 across 1,011 Saccharomyces cerevisiae isolates. *Nature* 2018; **556**: 339–344.
563 10. Bigey F, Segond D, Friedrich A, Guezenec S, Bourgais A, Huyghe L, et al. Evidence for Two
564 Main Domestication Trajectories in Saccharomyces cerevisiae Linked to Distinct Bread-Making
565 Processes. *Current Biology* 2020; S0960982220316912.
566 11. Arranz-Otaegui A, Carretero LG, Ramsey MN, Fuller DQ, Richter T. Archaeobotanical
567 evidence reveals the origins of bread 14,400 years ago in northeastern Jordan. *PNAS* 2018; **115**: 7925–
568 7930.
569 12. Samuel D. Bread in archaeology. *Civilisations Revue internationale d'anthropologie et de*
570 *sciences humaines* 2002; 27–36.
571 13. Roussel P, Chiron H. Les pains français: évolution, qualité, production. 2005. Maé-Erti.
572 14. Decock P, Cappelle S. Bread technology and sourdough technology. *Trends in Food Science*
573 *& Technology* 2005; **16**: 113–120.
574 15. Van Kerrebroeck S, Maes D, De Vuyst L. Sourdoughs as a function of their species diversity
575 and process conditions, a meta-analysis. *Trends in Food Science & Technology* 2017; **68**: 152–159.
576 16. Arora K, Ameur H, Polo A, Di Cagno R, Rizzello CG, Gobbetti M. Thirty years of knowledge
577 on sourdough fermentation: A systematic review. *Trends in Food Science & Technology* 2021; **108**:
578 71–83.
579 17. Landis EA, Oliverio AM, McKenney EA, Nichols LM, Kfoury N, Biango-Daniels M, et al.
580 The diversity and function of sourdough starter microbiomes. *eLife* 2021; **10**: e61644.
581 18. Demeulenaere É, Rivière P, Hyacinthe A, Baltassat R, Baltazar S, Gascuel J-S, et al. La
582 sélection participative à l'épreuve du changement d'échelle. À propos d'une collaboration entre
583 paysans sélectionneurs et généticiens de terrain. *Natures Sciences Societes* 2017; **Vol. 25**: 336–346.
584 19. Urien C, Legrand J, Montalent P, Casaregola S, Sicard D. Fungal Species Diversity in French
585 Bread Sourdoughs Made of Organic Wheat Flour. *Front Microbiol* 2019; **10**.
586 20. Michel E, Monfort C, Deffrasnes M, Guezenec S, Lhomme E, Barret M, et al.
587 Characterization of relative abundance of lactic acid bacteria species in French organic sourdough by
588 cultural, qPCR and MiSeq high-throughput sequencing methods. *International Journal of Food*

589 *Microbiology* 2016; **239**: 35–43.
590 21. White TJ, Bruns T, Lee S, Taylor J. 38 - AMPLIFICATION AND DIRECT SEQUENCING
591 OF FUNGAL RIBOSOMAL RNA GENES FOR PHYLOGENETICS. In: Innis MA, Gelfand DH,
592 Sninsky JJ, White TJ (eds). *PCR Protocols*. 1990. Academic Press, San Diego, pp 315–322.
593 22. Martin KJ, Rygiewicz PT. Fungal-specific PCR primers developed for analysis of the ITS
594 region of environmental DNA extracts. *BMC Microbiology* 2005; **5**: 28.
595 23. Peris D, Sylvester K, Libkind D, Gonçalves P, Sampaio J, Alexander WG, et al. Population
596 structure and reticulate evolution of Saccharomyces eubayanus and its lager-brewing hybrids. *Mol*
597 *Ecol* 2014; **23**: 2031–2045.
598 24. Nguyen H-V, Gaillardin C. Evolutionary relationships between the former species
599 Saccharomyces uvarum and the hybrids Saccharomyces bayanus and Saccharomyces pastorianus;
600 reinstatement of Saccharomyces uvarum (Beijerinck) as a distinct species. *FEMS Yeast Res* 2005; **5**:
601 471–483.
602 25. Libkind D, Hittinger C, of the … V-E. Microbe domestication and the identification of the
603 wild genetic stock of lager-brewing yeast. *Proceedings of the* … 2011.
604 26. Sampaio JP, Gonçalves P. Natural populations of Saccharomyces kudriavzevii in Portugal are
605 associated with oak bark and are sympatric with S. cerevisiae and S. paradoxus. *Appl Environ*
606 *Microbiol* 2008; **74**: 2144–2152.
607 27. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*
608 2016; **44**: D7–D19.
609 28. Weiss S, Samson F, Navarro D, Casaregola S. YeastIP: a database for identification and
610 phylogeny of Saccharomycotina yeasts. *FEMS Yeast Research* 2013; **13**: 117–125.
611 29. Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, et al. FROGS: Find,
612 Rapidly, OTUs with Galaxy Solution. *Bioinformatics* 2018; **34**: 1287–1294.
613 30. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome
614 assemblies. *Bioinformatics* 2011; **27**: 2957–2963.
615 31. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
616 *EMBnet.journal* 2011; **17**: 10–12.

617 32. Sickle 1.33 – Windowed Adaptive Trimming for Fastq files using Quality – My Biosoftware –
618 Bioinformatics Softwares Blog.
619 33. Mahé F, Rognes T, Quince C, Vargas C de, Dunthorn M. Swarm: robust and fast clustering
620 method for amplicon-based studies. *PeerJ* 2014; **2**: e593.
621 34. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool
622 for metagenomics. *PeerJ* 2016; **4**: e2584.
623 35. Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, et al. Towards a
624 unified paradigm for sequence-based identification of fungi. *Molecular Ecology* 2013; **22**: 5271–5277.
625 36. Bloem A, Rollero S, Seguinot P, Crépin L, Perez M, Picou C, et al. Workflow Based on the
626 Combination of Isotopic Tracer Experiments to Investigate Microbial Metabolism of Multiple Nutrient
627 Sources. *J Vis Exp* 2018.
628 37. Delobel P, Pradal M, Blondin B, Tesniere C. A 'fragile cell' sub-population revealed during
629 cytometric assessment of Saccharomyces cerevisiae viability in lipid-limited alcoholic fermentation.
630 *Lett Appl Microbiol* 2012; **55**: 338–344.
631 38. Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. *Journal of*
632 *Statistical Software* 2008; **25**: 1–18.
633 39. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and
634 Graphics of Microbiome Census Data. *PLOS ONE* 2013; **8**: e61217.
635 40. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial
636 communities. *Appl Environ Microbiol* 2005; **71**: 8228–8235.
637 41. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface
638 for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010; **27**: 221–224.
639 42. Chao A, Chiu C-H, Jost L. Phylogenetic diversity measures based on Hill numbers. *Philos*
640 *Trans R Soc Lond B Biol Sci* 2010; **365**: 3599–3609.
641 43. Jost L. Partitioning diversity into independent alpha and beta components. *Ecology* 2007; **88**:
642 2427–2439.
643 44. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful
644 Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*

1995; **57**: 289–300.

45. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 2017; **82**: 1–26.

46. Hijmans RJ, Karney (GeographicLib) C, Williams E, Vennes C. geosphere: Spherical Trigonometry. 2021.

47. Dray S, Dufour A-B. The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software* 2007; **22**: 1–20.

48. Pedersen HW Danielle Navarro, and Thomas Lin. Welcome | ggplot2.

49. Cheng J, Karambelkar B, Xie Y. Leaflet: Create interactive web maps with the javascript'leaflet'library. *R package version* 2018; **2**.

50. 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 Res Int* 2018; **106**: 974–981.

51. De Vuyst L, Van Kerrebroeck S, Leroy F. Microbial Ecology and Process Technology of Sourdough Fermentation. *Adv Appl Microbiol* 2017; **100**: 49–160.

52. 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.

53. 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 Microbiol* 2016; **53**: 41–50.

54. Goddard MR, Greig D. Saccharomyces cerevisiae: a nomadic yeast with no niche? *FEMS Yeast Res* 2015; **15**: fov009.

55. 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.

56. Wang X, Du H, Zhang Y, Xu Y. Environmental Microbiota Drives Microbial Succession and Metabolic Profiles during Chinese Liquor Fermentation. *Appl Environ Microbiol* 2018; **84**: e02369-

673 17.
674 57. Chiva R, Celador-Lera L, Uña JA, Jiménez-López A, Espinosa-Alcantud M, Mateos-
675 Horganero E, et al. Yeast Biodiversity in Fermented Doughs and Raw Cereal Matrices and the Study
676 of Technological Traits of Selected Strains Isolated in Spain. *Microorganisms* 2020; **9**: 47.
677 58. Reese AT, Madden AA, Joossens M, Lacaze G, Dunn RR. Influences of Ingredients and
678 Bakers on the Bacteria and Fungi in Sourdough Starters and Bread. *mSphere* 2020; **5**.
679 59. Syrokou MK, Themeli C, Paramithiotis S, Mataragas M, Bosnea L, Argyri AA, et al.
680 Microbial Ecology of Greek Wheat Sourdoughs, Identified by a Culture-Dependent and a Culture-
681 Independent Approach. *Foods* 2020; **9**: 1603.
682 60. Sugihara TF, Kline L, Miller MW. Microorganisms of the San Francisco sour dough bread
683 process. I. Yeasts responsible for the leavening action. *Appl Microbiol* 1971; **21**: 456–458.
684 61. Middelhoven WJ, Kurtzman CP, Vaughan-Martini A. Saccharomyces bulderi sp. nov., a yeast
685 that ferments gluconolactone. *Antonie Van Leeuwenhoek* 2000; **77**: 223–228.
686 62. Gori K, Bjørklund MK, Canibe N, Pedersen AØ, Jespersen L. Occurrence and identification
687 of yeast species in fermented liquid feed for piglets. *Microb Ecol* 2011; **61**: 146–153.
688 63. Comasio A, Verce M, Van Kerrebroeck S, De Vuyst L. Diverse Microbial Composition of
689 Sourdoughs From Different Origins. *Front Microbiol* 2020; **11**: 1212.
690 64. Gouliamova D, Dimitrov R. KAZACHSTANIA CHRYSOLINAE AND KAZACHSTANIA
691 BOZAE TWO NEW YEAST SPECIES OF THE GENUS KAZACHSTANIA. TRANSFER OF
692 FOUR KAZACHSTANIA SPECIES TO GRIGOROVIA GEN. NOV. AS NEW COMBINATIONS.
693 2020.
694 65. Jacques N, Sarilar V, Urien C, Lopes MR, Morais CG, Uetanabaro APT, et al. Three novel
695 ascomycetous yeast species of the Kazachstania clade, Kazachstania saulgeensis sp nov., Kazachstania
696 serrabonitensis sp nov and Kazachstania australis sp nov Reassignment of Candida humilis to
697 Kazachstania humilis f.a. comb. nov and Candida pseudohumilis to Kazachstania pseudohumilis f.a.
698 comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 2016; **66**: 5192--5200.
699 66. Sarilar V, Sterck L, Matsumoto S, Jacques N, Neuvéglise C, Tinsley CR, et al. Genome
700 sequence of the type strain CLIB 1764T (=CBS 14374T) of the yeast species Kazachstania saulgeensis

701 isolated from French organic sourdough. *Genomics Data* 2017; **13**: 41–43.
702 67. Minervini F, Lattanzi A, De Angelis M, Di Cagno R, Gobbetti M. Influence of Artisan
703 Bakery- or Laboratory-Propagated Sourdoughs on the Diversity of Lactic Acid Bacterium and Yeast
704 Microbiotas. *Appl Environ Microbiol* 2012; **78**: 5328–5340.
705 68. Salovaara H, Savolainen J. Yeast type isolated from Finnish sour rye dough starters. *Acta*
706 *Alimentaria Polonica* 1984; **10**.
707 69. Decimo M, Quattrini M, Ricci G, Fortina MG, Brasca M, Silvetti T, et al. Evaluation of
708 microbial consortia and chemical changes in spontaneous maize bran fermentation. *AMB Express*
709 2017; **7**: 205.
710 70. Fraberger V, Unger C, Kummer C, Domig KJ. Insights into microbial diversity of traditional
711 Austrian sourdough. *LWT* 2020; **127**: 109358.
712 71. Lhomme E, Lattanzi A, Dousset X, Minervini F, De Angelis M, Lacaze G, et al. Lactic acid
713 bacterium and yeast microbiotas of sixteen French traditional sourdoughs. *Int J Food Microbiol* 2015;
714 **215**: 161–170.
715 72. Jacques N, Sarilar V, Urien C, Lopes MR, Morais CG, Uetanabaro APT, et al. Three novel
716 ascomycetous yeast species of the Kazachstania clade, Kazachstania saulgeensis sp. nov.,
717 Kazachstaniaserrabonitensis sp. nov. and Kazachstania australis sp. nov. Reassignment of Candida
718 humilis to Kazachstania humilis fa comb. nov. and Candida pseudohumilis to Kazachstania
719 pseudohumilis fa comb. nov. *International journal of systematic and evolutionary microbiology* 2016;
720 **66**: 5192–5200.
721 73. Kurtzman CP. Phylogenetic circumscription of Saccharomyces, Kluyveromyces and other
722 members of the Saccharomycetaceae, and the proposal of the new genera Lachancea, Nakaseomyces,
723 Naumovia, Vanderwaltozyma and Zygotorulaspora. *FEMS yeast research* 2003; **4**: 233–245.
724 74. Boynton PJ, Greig D. The ecology and evolution of non-domesticated Saccharomyces species.
725 *Yeast* 2014; **31**: 449–462.
726 75. Minervini F, Lattanzi A, De Angelis M, Celano G, Gobbetti M. House microbiotas as sources
727 of lactic acid bacteria and yeasts in traditional Italian sourdoughs. *Food Microbiology* 2015; **52**: 66–
728 76.

729 76. Minervini F, Lattanzi A, Dinardo FR, De Angelis M, Gobbetti M. Wheat endophytic
730 lactobacilli drive the microbial and biochemical features of sourdoughs. *Food Microbiol* 2018; **70**:
731 162–171.
732

733 **Figure Legends**

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

747

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

750 A) Representation of bakers. Each point represents a bakery. The left area brings together baker with
751 "artisanal" practices and the right area bakers with "farmer" practices. Centroid of each group is
752 designed by a triangle. The dot's colors indicate the PCoA cluster of the sourdough fungal community
753 (see Figure 5). Black dots for group 1, white for group 2, grey for group 3. The fungal community of
754 the sourdough of baker 14 (in red) was not studied. B) Representation of the 20 first categories that
755 contributed the most to the MCA axes. The category is written next to the triangle. C) Distribution of
756 each variable for each bread making practices group. Only variables that mostly explained differences
757 between bread making practice groups are shown: use of commercial yeast, kneading method, chief
758 origin, kg of bread production per week, number of bread making per week, percent of water in dough,
759 number of back-sloppings before making bread, water origin, sourdough age and flour percentage in
760 dough.

761 **Figure 3.** Yeast species diversity was analyzed for 38 out of the 39 sourdoughs with both cultural and
762 metabarcoding methods.
763 Left: species were identified by traditional microbial isolation and identification using ITS sequencing.
764 Right: species were identified using ITS1 metabarcoding. The three most frequently-encountered
765 species are shown in contrasting colors (blue: *Saccharomyces cerevisiae*, red: *Kazachstania bulderi*,
766 yellow: *Kazachstania humilis*).
767
768 **Figure 4.** Distribution of yeast species diversity across French sourdoughs. Each bar represents the yeast
769 species diversity of one sourdough and is placed on the map where the baker is located.
770
771 **Figure 5.** Representation of sourdough depending on the weighted Unifrac distances between
772 microbial communities. The first line shows the clustering of sourdough according to their Unifrac
773 distance on a tree (left) or on a PCoA (right). Sourdough fungal community can be clustered in three
774 groups according to their weighted Unifrac distances. The lines below show the distribution of bread
775 making practices or the distribution of the dominant or most frequent species among fungal
776 community. Each line corresponds to a variable. On the left: a tree constructed from the Unifrac
777 distance matrix colored by groups of sourdoughs defined by the clustering analysis. In the center:
778 sourdoughs represented on the first 2 axes of the PCoA colored according to the group of sourdoughs.
779 On the right: the distribution of modalities of the variable for each group of sourdoughs (number of
780 each group on the x-axis).
781
782 **Figure 6**. Principal component analysis of 37 *K. humilis* and *K. bulderi* strains based on the
783 quantitative variation of maximum CO2 production (CO2max), fermentation latency phase (t1g),
784 maximum CO2 production rate (Vmax), time to reach the maximum production rate (tVmax), log of
785 population size and mortality at the end of the fermentation. The correlations between variables are
786 presented on the left while the figure on the right shows the projection of strains on the first two axes
787 representing 70.64% of the variation. The strains are colored according to their habitat of origin
788 (sourdough/non-sourdough). Their species is indicated by symbol.

789 **Figure 7**. Ratio between the sourdough strains mean and the non-sourdough strains mean values of
790 each quantitative variable measuring fermentation performance (t1g, tVmax, Vmax, CO2max) and
791 absolute fitness (Cellt27, Mortality). Confidence intervals are indicated by bars.
792

Figure 1

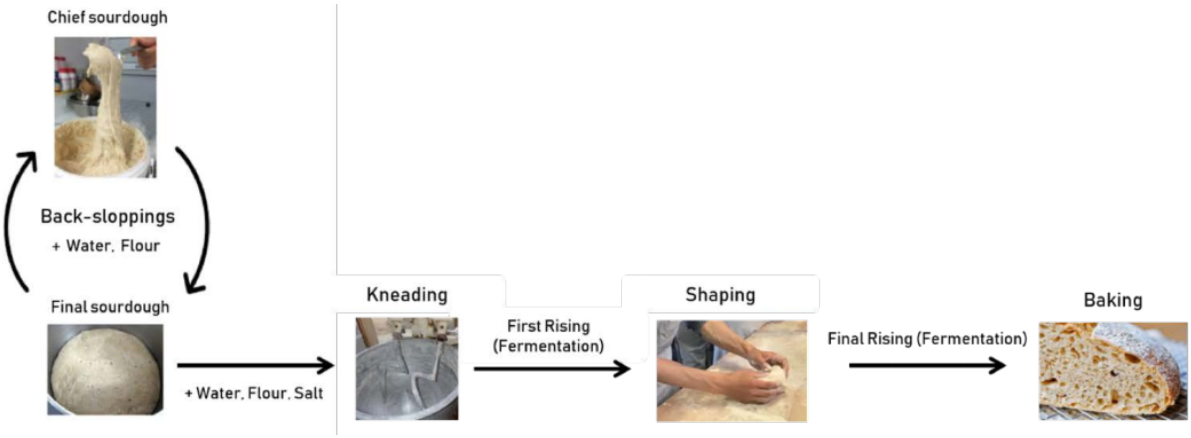

Chief sourdough
Back-sloppings
+ Water, Flour
Final sourdough
+ Water, Flour, Salt
Kneading
First Rising
(Fermentation)
Shaping
Final Rising (Fermentation)
Baking

Figure 2

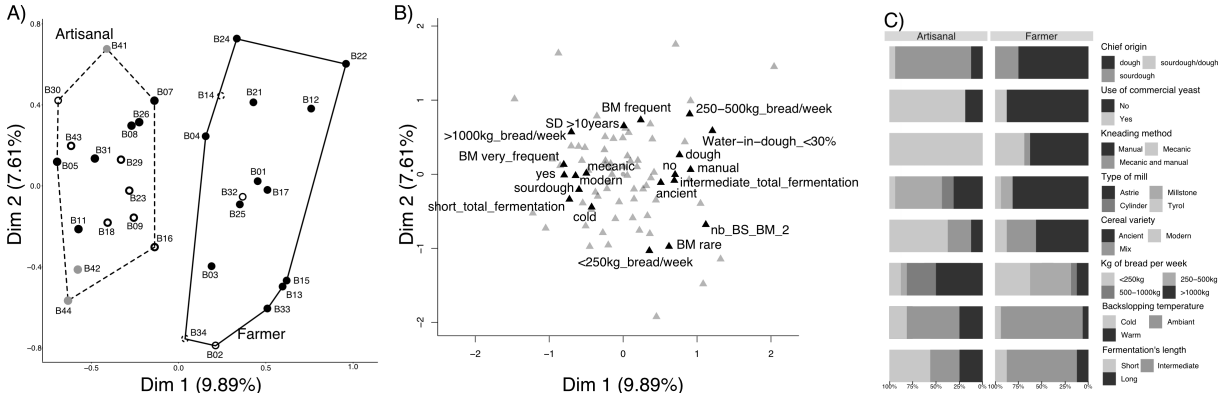

A)
Artisanal
Farmer
Dim 1 (9.89%)
Dim 2 (7.61%)
B)
BM frequent
250–500kg_bread/week
SD >10years
>1000kg_bread/week
Water–in–dough_<30%
BM very_frequent
dough
mecanic
no
manual
yes
modern
intermediate_total_fermentation
sourdough
ancient
short_total_fermentation
cold
nb_BS_BM_2
BM rare
<250kg_bread/week
Dim 1 (9.89%)
Dim 2 (7.61%)
C)
Artisanal
Farmer
Chief origin
dough
sourdough/dough
sourdough
Use of commercial yeast
No
Yes
Kneading method
Manual
Mecanic
Mecanic and manual
Type of mill
Astrie
Millstone
Cylinder
Tyrol
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

Figure 3

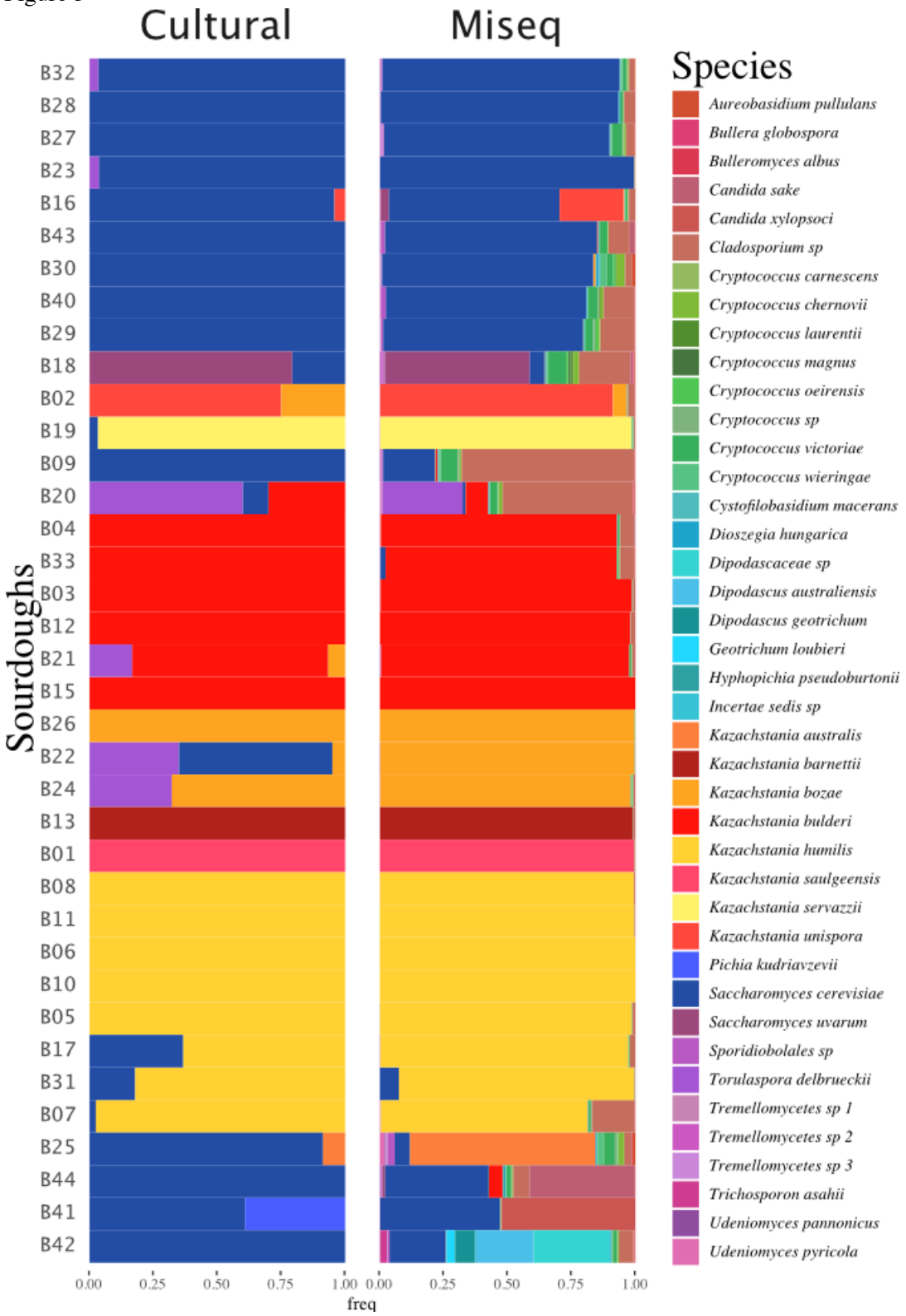

Cultural
Miseq
Species
Sourdoughs
B32
B28
B27
B23
B16
B43
B30
B40
B29
B18
B02
B19
B09
B20
B04
B33
B03
B12
B21
B15
B26
B22
B24
B13
B01
B08
B11
B06
B10
B05
B17
B31
B07
B25
B44
B41
B42
0.00
0.25
0.50
0.75
1.00
freq
Aureobasidium pullulans
Bullera globospora
Bulleromyces albus
Candida sake
Candida xylopsoci
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

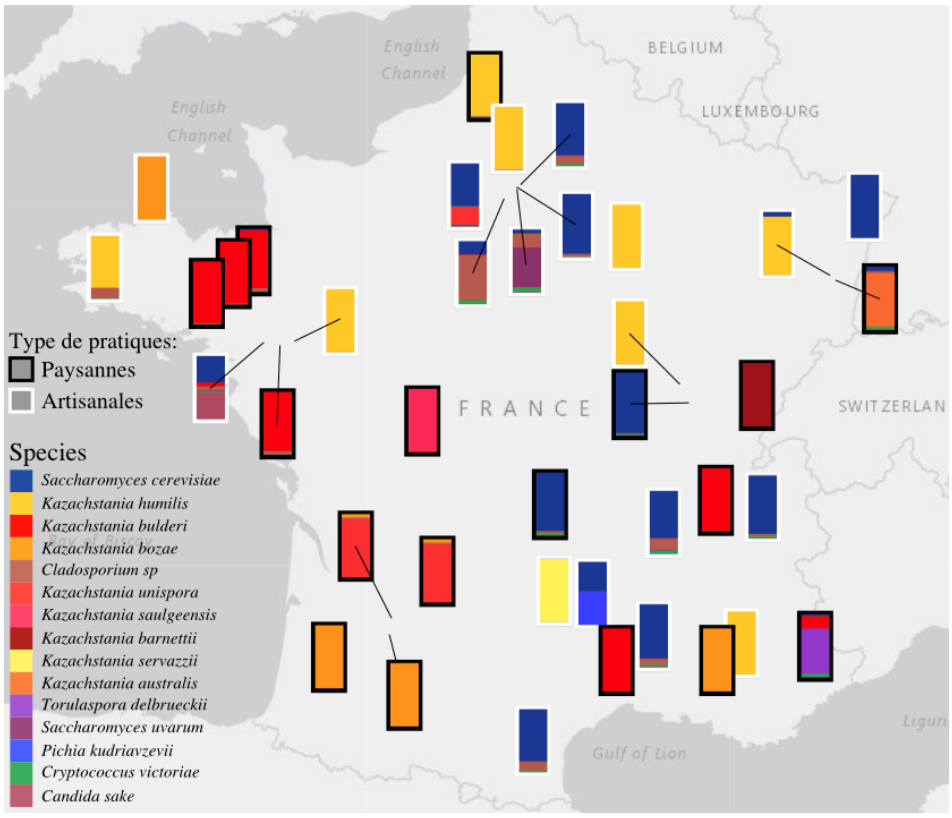

English Channel
English Channel
BELGIUM
LUXEMBOURG
FRANCE
SWITZERLAN
Gulf of Lion
Type de pratiques:
Paysannes
Artisanales
Species
Saccharomyces cerevisiae
Kazachstania humilis
Kazachstania bulderi
Kazachstania bozae
Cladosporium sp
Kazachstania unispora
Kazachstania saulgeensis
Kazachstania barnettii
Kazachstania servazzii
Kazachstania australis
Torulaspora delbrueckii
Saccharomyces uvarum
Pichia kudriavzevii
Cryptococcus victoriae
Candida sake

Figure 5

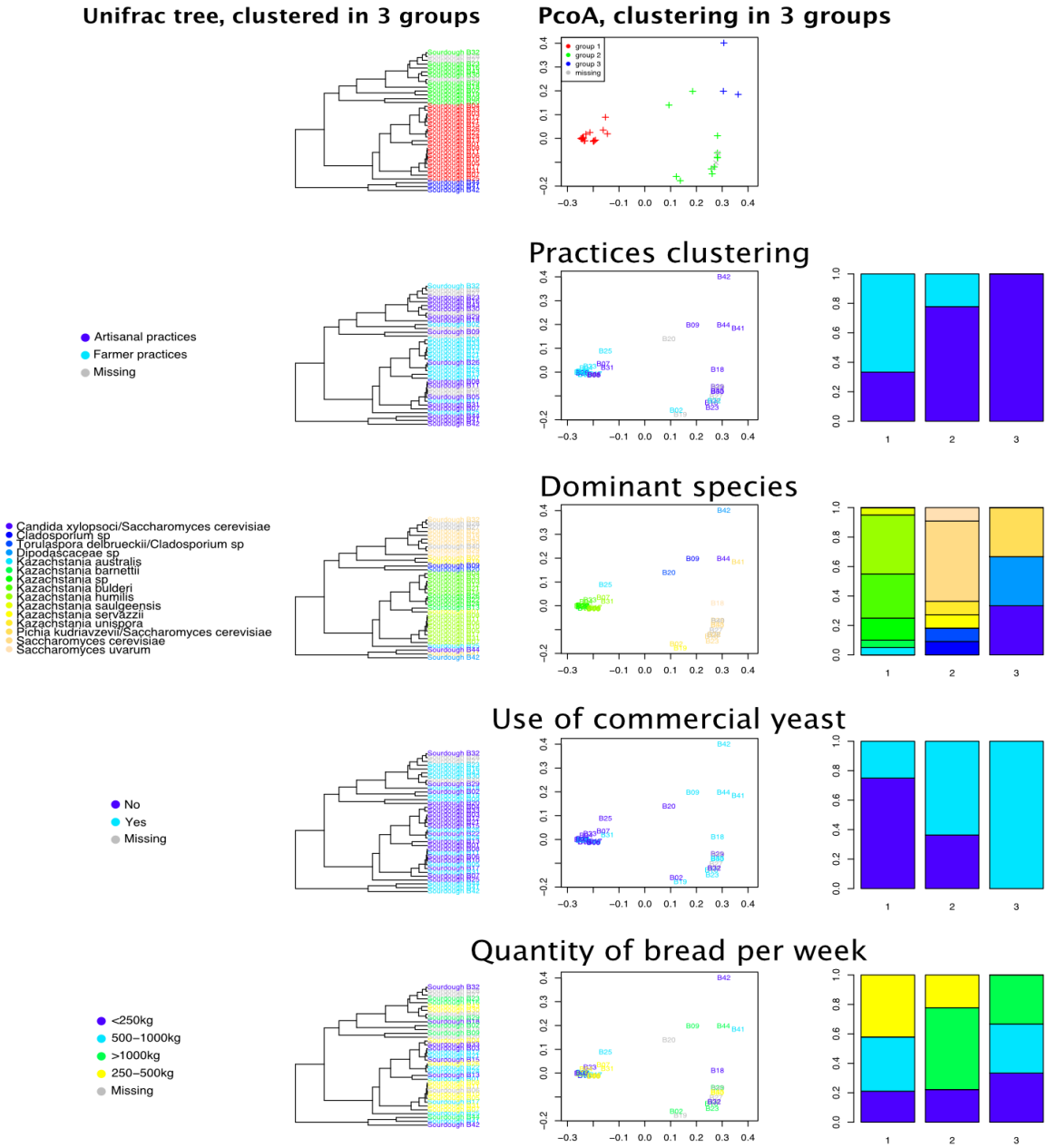

Unifrac tree, clustered in 3 groups
PcoA, clustering in 3 groups
group 1
group 2
group 3
missing
Practices clustering
Artisanal practices
Farmer practices
Missing
Dominant species
Candida xylopsoci/Saccharomyces cerevisiae
Cladosporium sp
Torulaspora delbrueckii/Cladosporium sp
Dipodascaceae sp
Kazachstania australis
Kazachstania barnettii
Kazachstania sp
Kazachstania bulderi
Kazachstania humilis
Kazachstania saulgeensis
Kazachstania servazzii
Kazachstania unispora
Pichia kudriavzevii/Saccharomyces cerevisiae
Saccharomyces cerevisiae
Saccharomyces uvarum
Use of commercial yeast
No
Yes
Missing
Quantity of bread per week
<250kg
500–1000kg
>1000kg
250–500kg
Missing

Figure 6

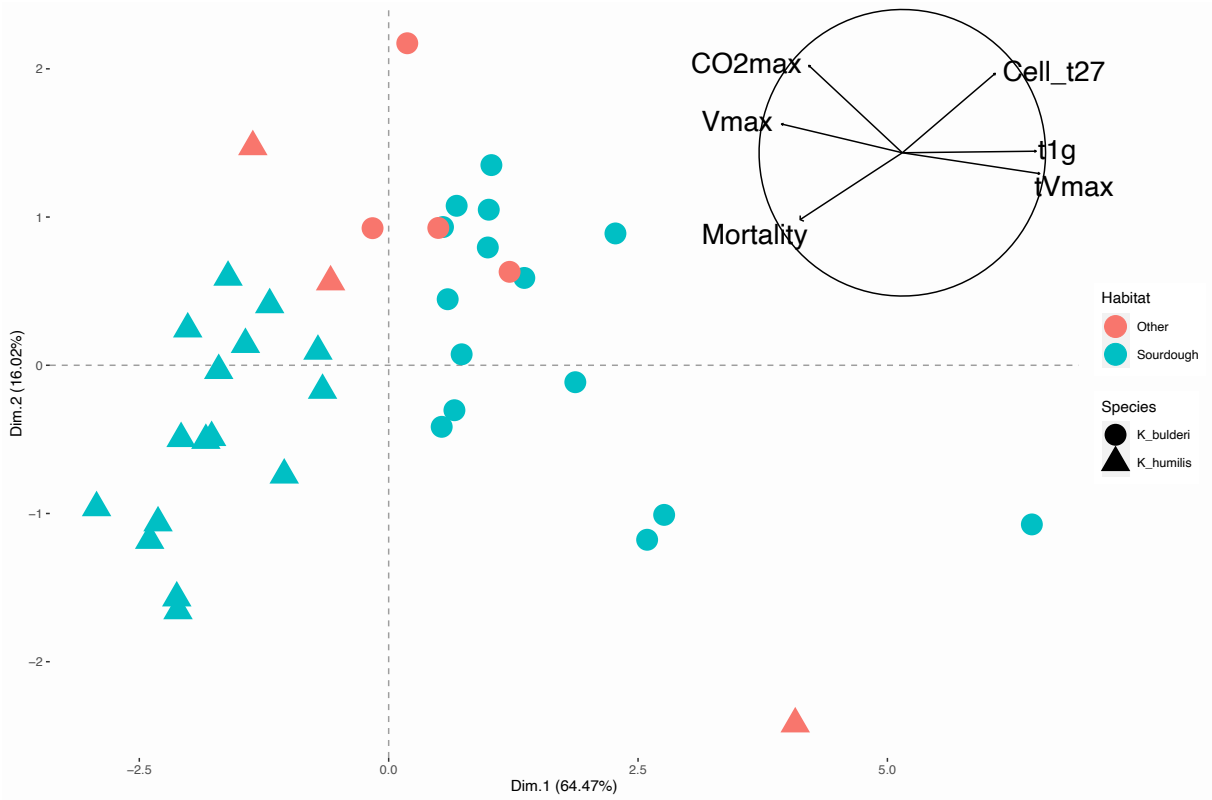
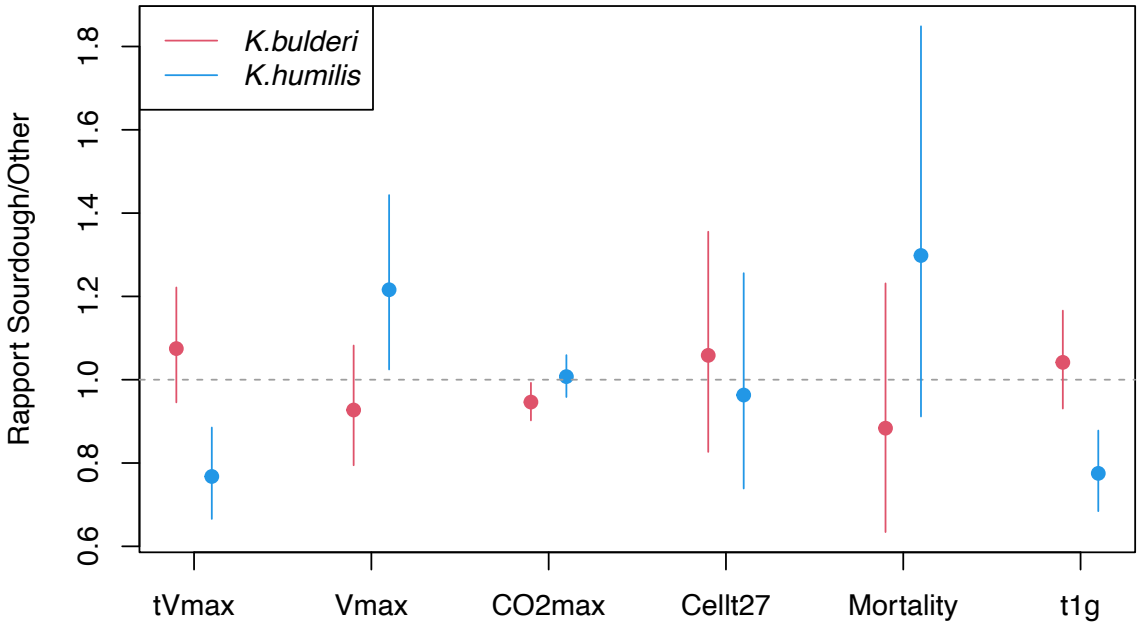

CO2max
Cell_t27
Vmax
t1g
tVmax
Mortality
Habitat
Other
Sourdough
Species
K_bulderi
K_humilis
Dim.2 (16.02%)
Dim.1 (64.47%)
2
1
0
−1
−2
−2.5
0.0
2.5
5.0

Figure 7


K.bulderi
K.humilis
Rapport Sourdough/Other
0.6
0.8
1.0
1.2
1.4
1.6
1.8
tVmax
Vmax
CO2max
Cellt27
Mortality
t1g