Title: Artisanal and farmers 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 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 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 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 phenotypic divergence between sourdough and non-sourdough strains 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 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 starters. This selection led to a reduction in species diversity 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 roqueforti* 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 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 that 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

83 medium. However, industrial and sourdough strains of *S. cerevisiae* differ genetically and
84 phenotypically, indicating that sourdough use contributes to the conservation of bread related *S.*
85 *cerevisiae* lineages [10].

86 Beyond *S. cerevisiae*, sourdough can also host other yeast species. To date, more than 40 yeast
87 species have been detected in sourdough [1, 15, 16]. The most frequently encountered species are
88 *Wickerhamomyces anomalus* and *Kazachstania humilis*. Several other species in the genus
89 *Kazachstania* (*Kazachstania barnettii*, *Kazachstania exigua*, *Kazachstania bulderi*, *Kazachstania*
90 *unispora, Kazachstania servazzii*) as well as several species in the polyphyletic genus Pichia have also
91 been recurrently detected. The factors that determine the distribution of these species are still unknown.
92 A recent large-scale study of 500 sourdoughs from four continents found no effect of geography or
93 factors related to bread-making practices such as age of sourdough, storage location, feeding frequency,
94 grain intake, house characteristics [17]. However, most of the sourdoughs in this study were made by
95 private citizens who probably did not maintain the sourdough microbial community in the same way as
96 a professional baker. To our knowledge, no studies have been conducted to date to investigate the effect
97 of bakers' bread-making practices on sourdough yeast community composition.

98 In France, sourdough breads are made both by bakers and farmers who also grow and mill their
99 own wheat. The number of farmers-bakers has increased in the 2000s with two motives: to grow wheat
100 varieties meeting their needs and to assert their independence from industry [18]. Although farmer-
101 bakers are less numerous than bakers, they participate in the renewed interest in local wheat varieties
102 and artisanal know-how, which may contribute to the conservation of both socio-cultural diversity and
103 microbial diversity.

104 Here, we used a participatory research approach involving psycho-sociologists, biologists, bio-
105 statisticians, bakers and farmers-bakers to study whether and how bakers and farmers-bakers contribute
106 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

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

136 **Yeast species identification**

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

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

149 The ITS1 region was targeted with the PCR primers ITS1-F (5′- CTTGGTCATTTAGAGGAAGTAA -
150 3′) and ITS2 (5’-GCTGCGTTCTTCATCGATGC-3’).

151 The sequencing run was performed with MiSeq Reagent Kit v3. 2015 [20]. Sequences were analyzed
152 through FROGS “Find Rapidly OTU with Galaxy Solution” [29] and home-made pipelines. Overlapped
153 reads were merged with Flash [30] with a minimum overlap of 10 nucleotides, a maximum overlap of
154 300 nucleotides and a maximum mismatch density of 0.1. Primers were removed with Cutadapt [31]
155 and data were cleaned with Sickle with quality-threshold and length-threshold equal to 20 [32]. Reads

156 were clustered with Swarm (d=3) [33] and chimeras deleted with VSEARCH [34]. Sequences were then
157 filtered on minimum abundance of 0.005% of all sequences. From the OTU's abundance table and for
158 each OTU, the taxonomic affiliation using UNITE Version 7.1, Release 2016-11-20 [35], YeastIP [28]
159 and our own databases [19] was obtained by blasting OTUs representative sequences against each
160 database.

161 **Phenotypic analysis of yeast strains**

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

177 **Data analyses**

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

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

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

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

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

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

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

232 (sourdough or non-sourdough) had an impact on strain phenotype, each log-transformed quantitative
233 variable was analyzed separately using a mixed linear model as described below. The experimental
234 design was unbalanced between the two blocks with very few non-sourdough strains in one of the two
235 blocks. Therefore, for each phenotype variable, we first estimated the block effect with a subset of 8
236 strains cultivated in both blocks using a linear model with two fixed effects : the strain and the block.
237 Additive models were used as the interactions terms were not significantly different from zero after
238 adjusting p-values with the Benjamini-Hochberg method. Second, each phenotype variable was
239 corrected for the block effect and analyzed with the mixed effect model: $Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_{ij} +$
240 $Z_k + \epsilon_{ijkl}$where Y_{ijkl}represents the log-transformed phenotype variable corrected for the block effect
241 for the strain *k*, from species *i (i=1,2)*, sampled in environment *j (j=1,2)*, observed for replicate *l*. μ
242 represents the mean of the phenotype variable, $\alpha_i$the additive effect of species i, $\beta_j$the additive effect of
243 environment *j*, and γ_{ij}their interaction. $Z_k$represents the gaussian random effect of strain *k* with $Z_k \sim$
244 $N(0, \sigma_s^2)$and $\epsilon_{ijkl}$the gaussian residuals with $\epsilon_{ijkl} \sim N(0, \sigma^2)$. For each species i, the impact of the
245 environment was quantified using the contrast $\Delta_i = \beta_S + \gamma_{i,S} - \beta_{NS} - \gamma_{i,NS}$ with NS standing for "non-
246 sourdough" and "S" for '"sourdough" and tests ($H_0 = \Delta_i = 0, H_1: \Delta_i \neq 0$) were performed and p-values
247 were adjusted using the Benjamini-Hochberg method. As log-transformed data were analyzed, the
248 exponential of this contrast can be interpreted as the ratio between the sourdough mean and the non-
249 sourdough mean. Confidence intervals and tests were performed using the doBy R-package.

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

253

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 (3 groups), observation during break-making workshops (2 workshops), and an online/phone survey (36 bakers). The general process of sourdough bread making is presented in Figure 1. We analyzed 28 variables around this general bread making process, 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" practices group, includes 6 bakers and 11 farmers-bakers using the following practices: low bread production (<500 kg per week, 81% of the bakers of the "farmer-like" group), use of ancient wheat populations (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" practices group consists of 12 bakers and 4 farmer-bakers having more intensive practices, characterized by a large bread production (>500 kg per week, 81%), mechanical kneading (100%), use of modern wheat varieties (63%), working at ambient temperature (56%), using commercial yeast starters in addition to sourdough for bread making or using commercial yeast starters for pastries and buns making (81%). In this second group, bakers tend to make their chief sourdough from a final sourdough.

277 **Composition of sourdough fungal communities**

278 Sourdough is a mix of flour and water naturally fermented by bacteria and yeasts. Sourdough yeast
279 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
280 per gram, as commonly found in sourdoughs from all over the world [1, 16, 50, 51]. We isolated 20 to
281 40 yeast strains from each sourdough by picking colonies randomly and identified species using ITS
282 sequence as well as other barcodes when needed (see M&M section). Among the 39 collected
283 sourdoughs, one (sourdough B14) did not give any colony in the laboratory, suggesting that his
284 sourdough microbiota was no longer alive. A total of 1216 strains were characterized from the other 38
285 sourdoughs. In addition, we developed an ITS1 meta-barcoding MiSeq sequencing method on
286 sourdough (see sup M&M). After filtering 5,360,620 raw ITS1 sequences for quality, abundance
287 (0.005%) and chimera, 3,542,801 sequences were further analyzed. Overall, the sequences clustered in
288 113 OTUs. The number of reads per sourdough ranged from 8421 to 194 557. Therefore, we carried out
289 our analysis on the rarefied matrix. Among all OTUs, 64 were identified as non-yeast, including 10
290 assigned to the order *Triticodae* (especially to the species *Triticum aestivum*), 50 assigned to assigned
291 to genera that include plant pathogen species together with saprobic ones, such as *Alternaria*, *Aspergillus*
292 or *Fusarium*, *Gibberella*, while 4 OTUs remained unidentified. Among the 40 yeast or yeast-like OTUs,
293 96% of total reads were assigned to the phylum *Ascomycota*, 87.5% to the order *Saccharomycetales* and
294 85.7% to the family *Saccharomycetaceae*. Only 4% of the total reads were assigned to the phylum
295 *Basidiomycota*. Overall, three OTUs assigned to the species *K. humilis*, *K. bulderi* and *S. cerevisiae*
296 represented 20.3%, 15.5% and 24.1% respectively of the total number of reads and 28.1%, 23.7% and
297 18.2% respectively of the number of reads identified as yeast species (Figure 3).

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

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

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

312 All sourdoughs but two had a dominant yeast species with a relative abundance over 50% and
313 many species with a lower relative abundance (Figure 3). Within sourdoughs, fungal species richness
314 ranged from 10 to 33, with a 23.5 median (Table S3). The effective number of species per sourdough
315 calculated from the Shannon diversity index ranged from 1 to 6.8 (Table S3), with 70% of sourdoughs
316 having an index below 2 (Table S3). Between-sourdough species differences were analyzed using
317 weighted Unifrac distances, computed from a phylogenetic tree built from the distances between OTUs
318 using *Sporidiobolales species* as root (Figure S2). Unifrac distances computed with four differently
319 rooted trees were highly positively correlated (Figure S3). Unifrac distances between sourdoughs ranged
320 from 0.0005 and 0.71, with a median of 0.49 and a mean of 0.52. The clustering of sourdoughs according
321 to their Unifrac distances is show Figure 5. There was no significant correlation between the Unifrac
322 distances and geographical distances between sourdoughs (Mantel test, P=0.35).

323 We then analyzed specifically the distribution of yeast species as yeast, together with lactic acid
324 bacteria, are the main functional player in a sourdough ecosystem and for bread quality. Over the 40
325 yeast species detected in the 38 sourdoughs, 12 had a relative abundance over 50% in at least one
326 sourdough, four had a relative abundance between 20% and 50% and 24 had a relative abundance below
327 10%. All dominant species (relative abundance over 50%) were fermentative yeast species, except in
328 one sourdough that had a C*ladosporium* species. We found all the sourdough yeast genera

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

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

333 *Saccharomyces cerevisiae* was found in 53% of all sourdoughs (95% confidence intervals=36%
334 - 69%) but was dominant (relative abundance over 50%) in only 24% (95% confidence intervals=11%
335 - 40%) (Figure 3). In two cases, *S. cerevisiae* co-occurred with another yeast species at similar relative
336 abundance. In the first case, *S. cerevisiae* was present at a relative abundance of 40% with *Candida sake*
337 at a 41% relative abundance. In the second case, it was found at a relative abundance of 47% with *Pichia*
338 *kudriavzevii* at a relative abundance of 52%. In all the other cases, *S. cerevisiae* had a relative abundance
339 below 21% and was found with other dominant yeast species, such as *Kazachstania australis*, *K. humilis*,
340 *Saccharomyces uvarum* or *Torulaspora delbrueckii*. This suggests that *S. cerevisiae* did not displace
341 other species and can indeed be out-competed by other species in sourdoughs.

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

343 *Kazachstania* was the most represented yeast genus over all sourdoughs, when considering both the
344 number of reads over all sourdoughs and the number of detected species. Indeed, this genus represented
345 57% of the total number of reads while *Saccharomyces* represented 26% of the total number of reads.
346 In addition, eight species of the *Kazachstania* genus were found in sourdough, while the *Saccharomyces*
347 genus was represented by two species (*S. uvarum* and *S. cerevisiae*) (Figure 3). The *Kazachstania* genus
348 is one of the closest genetically related genus to Saccharomyces and contained Crabtree positive yeasts,
349 able to ferment glucose even when oxygen is present if the amount of sugar is sufficient (Hagman &
350 Piskur 2015). *Kazachstania* species dominated in 54% (95% confidence intervals=36%-69%) of
351 sourdoughs while *Saccharomyces* species dominated in 27% only (95% confidence intervals=13%-
352 43%). *Kazachstania humilis*, followed by *K. bulderi* were the most commonly dominant *Kazachstania*
353 species, and found in respectively 21% (95% confidence intervals=10%-37%) and 15% of sourdoughs

354 (95% confidence intervals=6%-31%) (Figure 3). A recently described *Kazachstania* species,
355 *Kazachstania bozae*, was also identified in five sourdoughs (4.5%-29%) and found dominant in three
356 (1.7%-22%) [64]. Strains of this species were closely related to a strain previously isolated from boza,
357 a Bulgarian fermented drink, as estimated with ITS and LSU (D1D2) barcodes (Source: NCBI,
358 GenBank: KC118125.1 and KX369579.1). In addition, *Kazachstania saulgeensis,* a recently described
359 species [65, 66], was dominant in one sourdough (0.07%-14%). *Kazachstania unispora* and *K. servazzi*
360 which had previously been detected in sourdough were also found [17, 17, 53, 57, 58, 63, 67–71].
361 Finally, some *Kazachstania* species were detected for the first time as dominant in sourdoughs, whereas
362 they had been previously found in other environments, like soil (*K. australis*), sauerkraut (*K. barnettii*)
363 [72–74]. None of the previous studies on sourdough have evidenced as many *Kazachstania* species in
364 sourdough.

365 **The composition of sourdough fungal communities was associated with differences in bread-**
366 **making practices**

367 We tested whether sourdough fungal community beta diversity could be explained by bread-making
368 practices. To do so, we performed univariate PERMANOVA analysis on the 30 bakers with less than 8
369 missing values for the 29 bread-making practices variables (Table S4). The univariate analysis revealed
370 that the weighted Unifrac distance is structured according to the use of commercial yeast in bakery
371 ($P<0.05$). It also varied significantly with sourdough age, chief sourdough origin (dough, sourdough or
372 both), the quantity of bread produced per week, the milling method (cylinder, millstone, Astrie, Tyrol),
373 the type of wheat variety (ancient, modern or a mix thereof) and the fermentation duration. However,
374 after FDR correction for taking into account multiple testing, none of these variables significantly
375 explained Unifrac distances.

376 In order to understand further the relationship between sourdough fungal community composition
377 and bread-making practices, we clustered sourdoughs according to their fungal community composition,
378 on the basis of the PCoA of their weighted Unifrac distances. Then, we tested the link between the
379 fungal community group and the bread making practice group (farmers/artisanal practices group) as well

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

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

405

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

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

414 To investigate the relation between sourdough fungal communities and metabolic compounds, we
415 quantified sourdough hydration, yeast density, bacteria density, sourdough pH, Total Titrable Acidity
416 (TTA), sourdough concentration in seven sugars (maltose, glucose, fructose, raffinose, arabinose,
417 mannose, xylose), four alcohols (glycerol, ethanol, mannitol, meso-erythtritol), six acids (lactate,
418 acetate, glutarate, pyruvate, malate, succinate) and calculated the fermentative quotient (lactate over
419 acetate ratio). For each variable, there was a wide range of variation (Table S5). The principal component
420 analysis based on all variables showed no evidence of sourdough grouping (Figure S4). As expected in
421 fermentation, yeast density was positively correlated to ethanol ($r=0.74$, $P<0.001$), glycerol ($r=0.67$,
422 $P<0.001$), and acetate ($r=0.6$, $P<0.001$) concentration. However, it was not significantly correlated to
423 sugar concentrations. This might be explained by the co-occurrence of bacteria which have their own
424 metabolism and interact by competition and/or cross feeding with sourdough yeasts.

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

431 In addition, we tested whether variations in quantitative variables were associated with the fungal
432 community groups (Table S5). Group 3 microbial community sourdoughs (defined by PCoA clustering
433 on Unifrac distance, see below), which contains *S. cerevisiae* in co-dominance with a second yeast
434 species (*Candida sake*, *Pichia kudriavzevii* or a *Dipodascus* species), had a significantly higher mean
435 pH (mean pH_{group3}=4.2 against pH_{group1}=3.8, Tukey Contrasts, P<0.001), lower TTA (mean TTA_{group3}=7.7
436 against TTA_{group1}=17.1, Tukey Contrasts, P=0.002), and a higher maltose concentration (mean Maltose
437 $_{group3}$=52.8 mg/gr of sourdough against Maltose $_{group1}$=24.1 mg/gr of sourdough, Tukey Contrasts
438 P=0.002) than group1, having a *Kazachstania* dominant species. Compared to group 2 having in most
439 cases *S. cerevisiae* as dominant species, it also had higher pH (pH_{group2}=3.9, Tukey Contrasts, P=0.003),
440 and higher maltose concentration (Maltose $_{group2}$=23.7, Tukey contrast, P=0.003). These data may reflect
441 a lower fermentative activity for group 3 fungal community having two co-dominant species, and/or a
442 negative interaction effect of group3 fungal community on the activity of lactic acid bacteria (LAB),
443 which are the main producers of sourdough acids. Previous studies on the bacteria content of the same
444 sourdoughs showed that *L. sanfranciscensis* was most generally the dominant species, although *L.*
445 *heilongjiangensis, L. curvatus* or *L. brevis* were also found as dominant species [20, 52, 53]. We found
446 no significant correlation between LAB and yeast densities (r= -0.15, p= 0.45, Figure S4) but the link
447 between fungal and bacteria community might be species and strains dependent. Additional studies on
448 the interactions between fungal and bacterial communities need to be performed to better understand
449 how they may drive sourdough acidity and sugar content.

450 Finally, we analyzed whether the variations of each quantitative variable was associated with the
451 dominant yeast species. We only considered the 26 sourdoughs having either *S. cerevisiae* (9
452 sourdoughs), *K. humilis* (8 sourdoughs), *K. bulderi* (6 sourdoughs) or *K. bozae* (3 sourdoughs) as
453 dominant species, since the other yeast species were found dominant only once. The differences in
454 dominant species was not significantly associated to variation in sourdough sugar, acids or alcohol
455 concentration. However, on average, sourdoughs dominated by *K. bulderi* were more hydrated (63%
456 water content in average) than sourdoughs dominated by *K. humilis*, *K. bozae*, and *S. cerevisiae*, having
457 respectively 49%, 47%, 53 % water content in average (P<0.001 for the 3 Tukey Contrasts).

458 *Kazachstania bulderi* was found to be dominant only in sourdoughs made using farmers' practices, a
459 bread making practice group that was also found to be associated with more hydrated sourdoughs.
460 Additional experiments should be carried out to test whether this species has indeed a better fitness in
461 more hydrated sourdoughs or whether its presence in more hydrated sourdoughs is related to covariation
462 with other farmer practices.

463 **Phenotypic signatures of domestication**

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

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

479 Statistical comparison of sourdough and non-sourdough strains of *K. bulderi* and *K. humilis* for each
480 phenotypic variable revealed phenotypic divergence for *K. humilis* but not for *K. bulderi*. While the *K.*
481 *bulderi* sourdough strains did not ferment significantly faster than the non-sourdough strains, the *K.*
482 *humilis* sourdough strains showed significantly higher Vmax, lower t1g, and lower tVmax than the non-

483 sourdough *K. humilis* strains (Figure 7, Table S6). On average, they started fermentation two hours
484 before the others and reached Vmax three hours before the others. In addition, their Vmax were on
485 average 34% higher than the others.

486 **Discussion**

487 Sourdough microbial diversity has been intensively studied worldwide. Despite a cultural and historical
488 interest on bread in France, French sourdough fungal diversity was only partly characterized before this
489 study [19, 53, 71]. A recent large-scale (> 500 starters) study of sourdough microbial diversity revealed
490 the fungal diversity that can be detected over the globe across home-made sourdoughs [17]. Comparison
491 of these findings with the diversity found in France showed that all the yeast species detected at a relative
492 abundance over 1% in this international collection of sourdoughs were detected in French sourdoughs
493 except the species, *Wickerhanomyces anomalus*, *Pichia membranifaciens*, *Naumovozyma castellii*, and
494 *Saccharomyces bayanus*. Inversely, French baker's sourdough harbored some yeast species that were
495 never found elsewhere, such as *K. bozae*, *K. australis*, *K. saulgeensis* [65].

496 Beyond the baker yeast species, *S. cerevisiae*, the most represented yeast genus in French sourdough
497 was *Kazachstania*. Eight *Kazachstania* species were found as dominant yeast species in at least one
498 French sourdough. This includes *K. humilis*, *K. bulderi*, *K. barnettii*, *K. unispora*, *K. servazzii*, *K. bozae*,
499 *K. australis*, *K. saulgeensis*. Three *Kazachstania* species (*K. exigua, K. lodderae, K. naganishi*) already
500 reported in sourdough were not found in our collection of French sourdough. *Kazachstania lodderae*
501 and *K. naganishi* are rarely found in sourdough. By contrast, *K. exigua* is a frequently cited sourdough
502 species in the literature. It has been previously found in France, Finland [77], Italy [78, 79], Denmark
503 [80], Ethiopia [81], USA [17]. It is also the first species to have been isolated from a sourdough (in San
504 Francisco, [29]). However, the attempt to study its genetic variation has been hampered by the fact that
505 it probably originated from hybridization between unknown yeast species [82]. To date, the genus
506 *Kazachstania* is composed of more than 40 species, of which 11 are present in the sourdough. It is
507 possible that an adaptive radiation linked to the adaptation to different sourdoughs or to different
508 anthropized niches has taken place, as it has been observed for example in cichlids during their

adaptation to different lakes. Indeed, five of the *Kazachstania* species present in the sourdough have so far only been detected in human-related niches. These are *K. saulgeensis*, *K. barnettii*, *K. bozae*, *K. bulderi*, *K. humilis*. These species are genetically closer to each other than they are to *K. australis*, *K. servazzii* and *K. unispora* which have also been found in nature and are grouped in another part of the *Kazachstania* pylogenetic tree. Genomic analysis of these species would shed light on their evolution. So far, eight of the 11 *Kazachstania* species found in the sourdough have at least one genome assembly available in public databases, including *K. saulgeensis* and *K. barnettii* [44, 45], and the assemblies of *K. bulderi* and *K. humilis*, which were recently published in public databases (Bio project: PRJEB44438). The genomic and phenomic analysis of the large collection of *Kazachstania* strains obtained by our study, together with the world collection of *Kazachstania* strains, may shed light on the radiation and domestication processes of these species.

We found that yeast community composition was partly related with 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. 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 of selection 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

535 selected an increased fermentation rate. Farmer-bakers typically store their sourdough for several days
536 during the week and therefore make a lower number of backslopping during the week. In addition, they
537 make bread with longer fermentation times than artisanal bakers. It is therefore possible that they did
538 not select to accelerate the speed of fermentation and instead let natural selection in the sourdough
539 environment act alone. Alternatively, the lack of phenotypic divergence between sourdough and non-
540 sourdough strains may reflect the limitation of our sampling. *Kazachstania bulderi* has been reported
541 for the first time in anaerobic maize silage in the Netherlands and in fermented liquid feed for piglets
542 [61, 62], more recently in French, Belgium and Spain sourdoughs [19, 53, 57, 63] but to our knowledge
543 was never found in wild environment. Here, we compare sourdough strains with strains coming from
544 ensilage and animal feed. It is unknown whether ensilage and animal feed strains are wild strains or feral
545 strains that have escaped from other domesticated environments. This may explain why we did not detect
546 any phenotypic divergence between sourdough and non-sourdough strains of *K. bulderi*. Other than
547 fermentation phenotypes, there was no evidence of fitness differences between sourdough and non-
548 sourdough strains of *K. humilis* and *K. bulderi* in sourdough mimicking media. Additional experiments
549 in real dough should be performed to further test the effect of natural selection in this environment.

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

558 To our knowledge, this is the first evidence of the influence of artisanal practices on taxonomic
559 diversity in microbial communities. On the other hand, several studies have shown that making
560 fermented products could lead to the selection of divergent phenotypes and genotypes. This is the case

561 of sourdough and industrial populations of the baker's yeast *S. cerevisiae* that diverge among themselves
562 and have a better fermentation performance than non-baker's strains. As for beer populations, sourdough
563 populations have acquired a better capacity to assimilate maltose, linked at least in part to an increase
564 in the number of copies of the genes involved in the assimilation of maltose. Several studies on wine
565 populations of *S. cerevisiae*, *Torulaspora delbrueckii* et *Lachancea thermotolerans* also showed that
566 wine populations are genetically differentiated from strains from other environments and present
567 beneficial phenotypes in grape must and for wine quality. The analysis of filamentous cheese fungi *P.*
568 *roqueforti* and *P. camemberti* populations also revealed cheese making genetically differentiated
569 populations. Interestingly, different genetic groups associated with different cheese making practices
570 were found. Strains of the blue cheese fungus, *P. roqueforti*, isolated from Roquefort cheese were more
571 diverse and were genetically and phenotypically different than strains used to make other blue cheeses
572 [83, 84, 85]. Two varieties of the white cheese making fungus, *P. camemberti*, with different phenotypic
573 features, were associated with different kind of cheese (camembert and brie). All together these studies
574 show that the diversity of practices used to make fermented products allows to maintain genetic,
575 phenotypic and taxonomic diversity.

576 However, fungal domestication also involved important bottleneck. For example, the low level of
577 genetic diversity found in blue-cheese P. roqueforti strains and P. camemberti strains revealed the risk
578 of diversity erosion in fermented product making [6, 7, 84]. This risk is associated with the massive use
579 of few industrial strains or the need to standardize products to meet the specifications of industrial
580 production or AOP. Here, we show that despite the recurrent use of S. cerevisiae as industrial starter
581 species in bakeries and homes, and the occurrence of this species in a wide range of habitats such as
582 soil, trees, and humans [9, 54], this species does not appear to have overwhelmingly colonized French
583 traditional sourdoughs (Figure 3, Figure 4). This result confirmed a recent analysis which revealed that
584 S. cerevisiae sourdough strains have a different evolutionary history than industrial strains [10]. The
585 dynamic of microbial species colonization and invasion in food environment remain largely unknown.
586 A recent study suggested that wild-type Penicillium can evolved in a few weeks into a domesticated
587 form. Additional experiments at the level of microbial community will shed light on the dynamic of

588 microbial community establishment in food production and on the ability of industrial strains to invade
589 food microbial community.

590

591 **Conclusion and perspectives**

592 In conclusion, a great diversity of bread-making practices and fungal community composition
593 was found in our sample of French sourdoughs. Surprisingly, the well-known baker's yeast
594 *Saccharomyces cerevisiae* was found dominant only in one fourth of the sampled sourdoughs. By
595 contrast, several species of the neighboring genus *Kazachstania* were detected at high frequency,
596 revealing a major role for this mostly unknown genus in the study of fungal domestication and in bread
597 making. Therefore, our results highlight the necessity of maintaining socio-cultural diversity to maintain
598 microbial diversity in food systems. These findings could not have been evidenced without the
599 collaboration of bakers and scientists, showing the importance of participatory research projects to gain
600 new insight into biodiversity preservation.

601

602 **Acknowledgements**

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

608

609 **References**

610 1. Carbonetto B, Ramsayer J, Nidelet T, Legrand J, Sicard D. Bakery yeasts, a new model for
611 studies in ecology and evolution. *Yeast* ; **0**.

612 2. Gibbons JG, Rinker DC. The genomics of microbial domestication in the fermented food
613 environment. *Current Opinion in Genetics & Development* 2015; **35**: 1–8.
614 3. Steensels J, Gallone B, Voordeckers K, Verstrepen KJ. Domestication of Industrial Microbes.
615 *Current Biology* 2019; **29**: R381–R393.
616 4. Gibbons JG, Salichos L, Slot JC, Rinker DC, McGary KL, King JG, et al. The evolutionary
617 imprint of domestication on genome variation and function of the filamentous fungus Aspergillus
618 oryzae. *Curr Biol* 2012; **22**: 1403–1409.
619 5. Watarai N, Yamamoto N, Sawada K, Yamada T. Evolution of Aspergillus oryzae before and
620 after domestication inferred by large-scale comparative genomic analysis. *DNA Research* 2019; **26**:
621 465–472.
622 6. Ropars J, Didiot E, Rodríguez de la Vega RC, Bennetot B, Coton M, Poirier E, et al.
623 Domestication of the Emblematic White Cheese-Making Fungus Penicillium camemberti and Its
624 Diversification into Two Varieties. *Curr Biol* 2020; **30**: 4441-4453.e4.
625 7. Ropars J, López-Villavicencio M, Snirc A, Lacoste S, Giraud T. Blue cheese-making has
626 shaped the population genetic structure of the mould Penicillium roqueforti. *PLOS ONE* 2017; **12**:
627 e0171387.
628 8. Legras J-L, Galeote V, Bigey F, Camarasa C, Marsit S, Nidelet T, et al. Adaptation of S.
629 cerevisiae to Fermented Food Environments Reveals Remarkable Genome Plasticity and the
630 Footprints of Domestication. *Mol Biol Evol* 2018; **35**: 1712–1727.
631 9. Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, et al. Genome evolution
632 across 1,011 Saccharomyces cerevisiae isolates. *Nature* 2018; **556**: 339–344.
633 10. Bigey F, Segond D, Friedrich A, Guezenec S, Bourgais A, Huyghe L, et al. Evidence for Two
634 Main Domestication Trajectories in Saccharomyces cerevisiae Linked to Distinct Bread-Making
635 Processes. *Current Biology* 2020; S0960982220316912.
636 11. Arranz-Otaegui A, Carretero LG, Ramsey MN, Fuller DQ, Richter T. Archaeobotanical
637 evidence reveals the origins of bread 14,400 years ago in northeastern Jordan. *PNAS* 2018; **115**: 7925–
638 7930.

12. Samuel D. Bread in archaeology. *Civilisations Revue internationale d'anthropologie et de sciences humaines* 2002; 27–36.

13. Roussel P, Chiron H. Les pains français: évolution, qualité, production. 2005. Maé-Erti.

14. Decock P, Cappelle S. Bread technology and sourdough technology. *Trends in Food Science & Technology* 2005; **16**: 113–120.

15. Van Kerrebroeck S, Maes D, De Vuyst L. Sourdoughs as a function of their species diversity and process conditions, a meta-analysis. *Trends in Food Science & Technology* 2017; **68**: 152–159.

16. Arora K, Ameur H, Polo A, Di Cagno R, Rizzello CG, Gobbetti M. Thirty years of knowledge on sourdough fermentation: A systematic review. *Trends in Food Science & Technology* 2021; **108**: 71–83.

17. Landis EA, Oliverio AM, McKenney EA, Nichols LM, Kfoury N, Biango-Daniels M, et al. The diversity and function of sourdough starter microbiomes. *eLife* 2021; **10**: e61644.

18. Demeulenaere É, Rivière P, Hyacinthe A, Baltassat R, Baltazar S, Gascuel J-S, et al. La sélection participative à l'épreuve du changement d'échelle. À propos d'une collaboration entre paysans sélectionneurs et généticiens de terrain. *Natures Sciences Societes* 2017; **Vol. 25**: 336–346.

19. Urien C, Legrand J, Montalent P, Casaregola S, Sicard D. Fungal Species Diversity in French Bread Sourdoughs Made of Organic Wheat Flour. *Front Microbiol* 2019; **10**.

20. Michel E, Monfort C, Deffrasnes M, Guezenec S, Lhomme E, Barret M, et al. Characterization of relative abundance of lactic acid bacteria species in French organic sourdough by cultural, qPCR and MiSeq high-throughput sequencing methods. *International Journal of Food Microbiology* 2016; **239**: 35–43.

21. White TJ, Bruns T, Lee S, Taylor J. 38 - AMPLIFICATION AND DIRECT SEQUENCING OF FUNGAL RIBOSOMAL RNA GENES FOR PHYLOGENETICS. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds). *PCR Protocols*. 1990. Academic Press, San Diego, pp 315–322.

22. Martin KJ, Rygiewicz PT. Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. *BMC Microbiology* 2005; **5**: 28.

23. Peris D, Sylvester K, Libkind D, Gonçalves P, Sampaio J, Alexander WG, et al. Population

666 structure and reticulate evolution of Saccharomyces eubayanus and its lager-brewing hybrids. *Mol*
667 *Ecol* 2014; **23**: 2031–2045.
668 24. Nguyen H-V, Gaillardin C. Evolutionary relationships between the former species
669 Saccharomyces uvarum and the hybrids Saccharomyces bayanus and Saccharomyces pastorianus;
670 reinstatement of Saccharomyces uvarum (Beijerinck) as a distinct species. *FEMS Yeast Res* 2005; **5**:
671 471–483.
672 25. Libkind D, Hittinger C, of the … V-E. Microbe domestication and the identification of the
673 wild genetic stock of lager-brewing yeast. *Proceedings of the* … 2011.
674 26. Sampaio JP, Gonçalves P. Natural populations of Saccharomyces kudriavzevii in Portugal are
675 associated with oak bark and are sympatric with S. cerevisiae and S. paradoxus. *Appl Environ*
676 *Microbiol* 2008; **74**: 2144–2152.
677 27. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*
678 2016; **44**: D7–D19.
679 28. Weiss S, Samson F, Navarro D, Casaregola S. YeastIP: a database for identification and
680 phylogeny of Saccharomycotina yeasts. *FEMS Yeast Research* 2013; **13**: 117–125.
681 29. Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, et al. FROGS: Find,
682 Rapidly, OTUs with Galaxy Solution. *Bioinformatics* 2018; **34**: 1287–1294.
683 30. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome
684 assemblies. *Bioinformatics* 2011; **27**: 2957–2963.
685 31. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
686 *EMBnet.journal* 2011; **17**: 10–12.
687 32. Sickle 1.33 – Windowed Adaptive Trimming for Fastq files using Quality – My Biosoftware –
688 Bioinformatics Softwares Blog.
689 33. Mahé F, Rognes T, Quince C, Vargas C de, Dunthorn M. Swarm: robust and fast clustering
690 method for amplicon-based studies. *PeerJ* 2014; **2**: e593.
691 34. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool
692 for metagenomics. *PeerJ* 2016; **4**: e2584.

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

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

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

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

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

40. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 2005; **71**: 8228–8235.

41. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010; **27**: 221–224.

42. Chao A, Chiu C-H, Jost L. Phylogenetic diversity measures based on Hill numbers. *Philos Trans R Soc Lond B Biol Sci* 2010; **365**: 3599–3609.

43. Jost L. Partitioning diversity into independent alpha and beta components. *Ecology* 2007; **88**: 2427–2439.

44. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 1995; **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.

720 47. Dray S, Dufour A-B. The ade4 Package: Implementing the Duality Diagram for Ecologists.
721 *Journal of Statistical Software* 2007; **22**: 1–20.
722 48. Pedersen HW Danielle Navarro, and Thomas Lin. Welcome | ggplot2.
723 49. Cheng J, Karambelkar B, Xie Y. Leaflet: Create interactive web maps with the
724 javascript'leaflet'library. *R package version* 2018; **2**.
725 50. Clément H, Prost C, Chiron H, Ducasse MB, Della Valle G, Courcoux P, et al. The effect of
726 organic wheat flour by-products on sourdough performances assessed by a multi-criteria approach.
727 *Food Res Int* 2018; **106**: 974–981.
728 51. De Vuyst L, Van Kerrebroeck S, Leroy F. Microbial Ecology and Process Technology of
729 Sourdough Fermentation. *Adv Appl Microbiol* 2017; **100**: 49–160.
730 52. Lhomme E, Orain S, Courcoux P, Onno B, Dousset X. The predominance of Lactobacillus
731 sanfranciscensis in French organic sourdoughs and its impact on related bread characteristics.
732 *International Journal of Food Microbiology* 2015; **213**: 40–48.
733 53. Lhomme E, Urien C, Legrand J, Dousset X, Onno B, Sicard D. Sourdough microbial
734 community dynamics: An analysis during French organic bread-making processes. *Food Microbiol*
735 2016; **53**: 41–50.
736 54. Goddard MR, Greig D. Saccharomyces cerevisiae: a nomadic yeast with no niche? *FEMS*
737 *Yeast Res* 2015; **15**: fov009.
738 55. Valmorri S, Tofalo R, Settanni L, Corsetti A, Suzzi G. Yeast microbiota associated with
739 spontaneous sourdough fermentations in the production of traditional wheat sourdough breads of the
740 Abruzzo region (Italy). *Antonie Van Leeuwenhoek* 2010; **97**: 119–129.
741 56. Wang X, Du H, Zhang Y, Xu Y. Environmental Microbiota Drives Microbial Succession and
742 Metabolic Profiles during Chinese Liquor Fermentation. *Appl Environ Microbiol* 2018; **84**: e02369-
743 17.
744 57. Chiva R, Celador-Lera L, Uña JA, Jiménez-López A, Espinosa-Alcantud M, Mateos-
745 Horganero E, et al. Yeast Biodiversity in Fermented Doughs and Raw Cereal Matrices and the Study
746 of Technological Traits of Selected Strains Isolated in Spain. *Microorganisms* 2020; **9**: 47.

58. Reese AT, Madden AA, Joossens M, Lacaze G, Dunn RR. Influences of Ingredients and Bakers on the Bacteria and Fungi in Sourdough Starters and Bread. *mSphere* 2020; **5**.

59. Syrokou MK, Themeli C, Paramithiotis S, Mataragas M, Bosnea L, Argyri AA, et al. Microbial Ecology of Greek Wheat Sourdoughs, Identified by a Culture-Dependent and a Culture-Independent Approach. *Foods* 2020; **9**: 1603.

60. Sugihara TF, Kline L, Miller MW. Microorganisms of the San Francisco sour dough bread process. I. Yeasts responsible for the leavening action. *Appl Microbiol* 1971; **21**: 456–458.

61. Middelhoven WJ, Kurtzman CP, Vaughan-Martini A. Saccharomyces bulderi sp. nov., a yeast that ferments gluconolactone. *Antonie Van Leeuwenhoek* 2000; **77**: 223–228.

62. Gori K, Bjørklund MK, Canibe N, Pedersen AØ, Jespersen L. Occurrence and identification of yeast species in fermented liquid feed for piglets. *Microb Ecol* 2011; **61**: 146–153.

63. Comasio A, Verce M, Van Kerrebroeck S, De Vuyst L. Diverse Microbial Composition of Sourdoughs From Different Origins. *Front Microbiol* 2020; **11**: 1212.

64. Gouliamova D, Dimitrov R. *Kazachstania chrysolinae* and *Kazachstania bozae* two new species of the genus *Kazachstania*. Transfer of four *Kazachstania* species to *Grigorovia* gen. nov. as new combinations. 2020.

65. 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 f.a. comb. nov and Candida pseudohumilis to Kazachstania pseudohumilis f.a. comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 2016; **66**: 5192--5200.

66. Sarilar V, Sterck L, Matsumoto S, Jacques N, Neuvéglise C, Tinsley CR, et al. Genome sequence of the type strain CLIB 1764T (=CBS 14374T) of the yeast species Kazachstania saulgeensis isolated from French organic sourdough. *Genomics Data* 2017; **13**: 41–43.

67. 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**: 5328–5340.

774 68. Salovaara H, Savolainen J. Yeast type isolated from Finnish sour rye dough starters. *Acta*
775 *Alimentaria Polonica* 1984; **10**.
776 69. Decimo M, Quattrini M, Ricci G, Fortina MG, Brasca M, Silvetti T, et al. Evaluation of
777 microbial consortia and chemical changes in spontaneous maize bran fermentation. *AMB Express*
778 2017; **7**: 205.
779 70. Fraberger V, Unger C, Kummer C, Domig KJ. Insights into microbial diversity of traditional
780 Austrian sourdough. *LWT* 2020; **127**: 109358.
781 71. Lhomme E, Lattanzi A, Dousset X, Minervini F, De Angelis M, Lacaze G, et al. Lactic acid
782 bacterium and yeast microbiotas of sixteen French traditional sourdoughs. *Int J Food Microbiol* 2015;
783 **215**: 161–170.
784 72. Jacques N, Sarilar V, Urien C, Lopes MR, Morais CG, Uetanabaro APT, et al. Three novel
785 ascomycetous yeast species of the Kazachstania clade, Kazachstania saulgeensis sp. nov.,
786 Kazachstaniaserrabonitensis sp. nov. and Kazachstania australis sp. nov. Reassignment of Candida
787 humilis to Kazachstania humilis fa comb. nov. and Candida pseudohumilis to Kazachstania
788 pseudohumilis fa comb. nov. *International journal of systematic and evolutionary microbiology* 2016;
789 **66**: 5192–5200.
790 73. Kurtzman CP. Phylogenetic circumscription of Saccharomyces, Kluyveromyces and other
791 members of the Saccharomycetaceae, and the proposal of the new genera Lachancea, Nakaseomyces,
792 Naumovia, Vanderwaltozyma and Zygotorulaspora. *FEMS yeast research* 2003; **4**: 233–245.
793 74. Boynton PJ, Greig D. The ecology and evolution of non-domesticated Saccharomyces species.
794 *Yeast* 2014; **31**: 449–462.
795 75. Minervini F, Lattanzi A, De Angelis M, Celano G, Gobbetti M. House microbiotas as sources
796 of lactic acid bacteria and yeasts in traditional Italian sourdoughs. *Food Microbiology* 2015; **52**: 66–76
797 76. Minervini F, Lattanzi A, Dinardo FR, De Angelis M, Gobbetti M. Wheat endophytic
798 lactobacilli drive the microbial and biochemical features of sourdoughs. *Food Microbiol* 2018; **70**:
799 162–171.
800 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.

825 **Figure Legends**

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

839

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

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

percent of water in dough, number of back-sloppings before making bread, water origin, sourdough age and flour percentage in dough. The categories of each of these variables are indicated on the right.

Figure 3. Yeast species diversity was analyzed for 38 out of the 39 sourdoughs with both cultural and metabarcoding methods. Left: species were identified by traditional microbial isolation and identification using ITS sequencing. Right: species were identified using ITS1 metabarcoding. The three most frequently-encountered species are shown in contrasting colors surrounded by black (blue: *Saccharomyces cerevisiae*, red: *Kazachstania bulderi*, yellow: *Kazachstania humilis*). The bread-making practice of the baker who supplied the sourdough is indicated on the left ("artisanal" in purple, "farmer" in light blue).

Figure 4. Distribution of yeast species diversity across French sourdoughs. Each bar represents the yeast species diversity of one sourdough and is placed on the map where the baker is located. Sourdoughs from "farmer" practice are surrounded in light blue and sourdoughs from "artisanal" practice are surrounded by purple.

Figure 5. Representation of sourdough depending on the weighted Unifrac distances between microbial communities. A) shows the clustering of sourdoughs according to their Unifrac distances on a tree (left) or on a PCoA (right). Sourdough fungal community can be clustered in three groups according to their weighted Unifrac distances. B to E) shows on the left: a tree constructed from the Unifrac distance matrix, in the center: sourdoughs represented on the first 2 axes of the PCoA and on the right: the distribution of modalities of a variable for each group of sourdoughs (number of each group on the x-axis). B) shows the distribution of bread-making practices among fungal community groups, C) shows the distribution of the dominant or most frequent species among fungal community (K: Kazachstania, C: Candida, T: Torulaspora, P: Pichia), D) shows the use of commercial *Saccharomyces cerevisiae* starter and E) shows the quantity of bread per week.

878 **Figure 6**. Principal component analysis of 37 *K. humilis* and *K. bulderi* strains based on the
879 quantitative variation of maximum CO2 production (CO2max), fermentation latency phase (t1g),
880 maximum CO2 production rate (Vmax), time to reach the maximum production rate (tVmax), log of
881 population size and mortality at the end of the fermentation. The correlations between variables are
882 presented on the left while the figure on the right shows the projection of strains on the first two axes
883 representing 70.64% of the variation. The strains are colored according to their habitat of origin
884 (sourdough/non-sourdough). Their species is indicated by symbol.
885
886 **Figure 7**. Ratio between the sourdough strains mean and the non-sourdough strains mean values of
887 each quantitative variable measuring fermentation performance: maximum CO2 production
888 (CO2max), fermentation latency phase (t1g), maximum CO2 production rate (Vmax), time to reach
889 the maximum production rate (tVmax), log of population size. Confidence intervals are indicated by
890 bars.

Figure 1

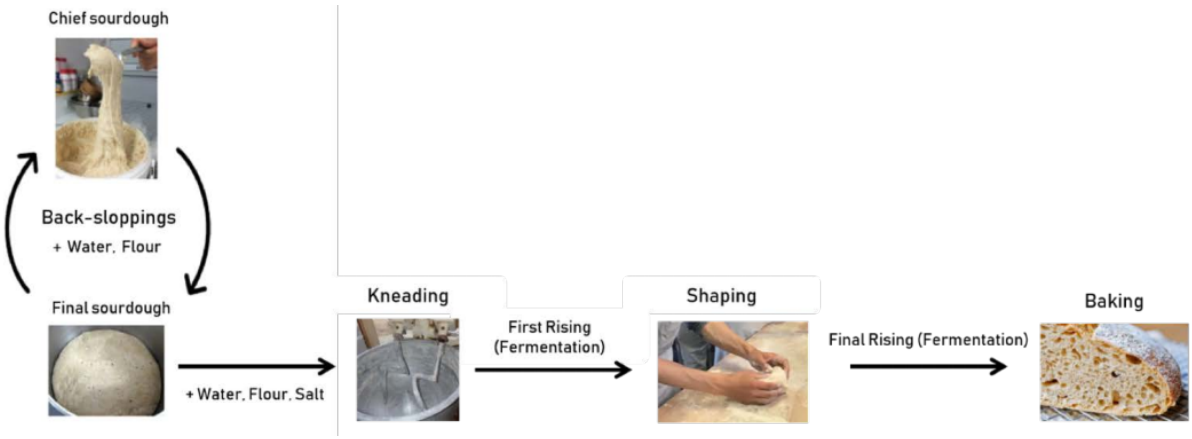

Chief sourdough
Back-sloppings
+ Water, Flour
Final sourdough
+ Water, Flour, Salt
Kneading
First Rising
(Fermentation)
Shaping
Final Rising (Fermentation)
Baking

Figure 2

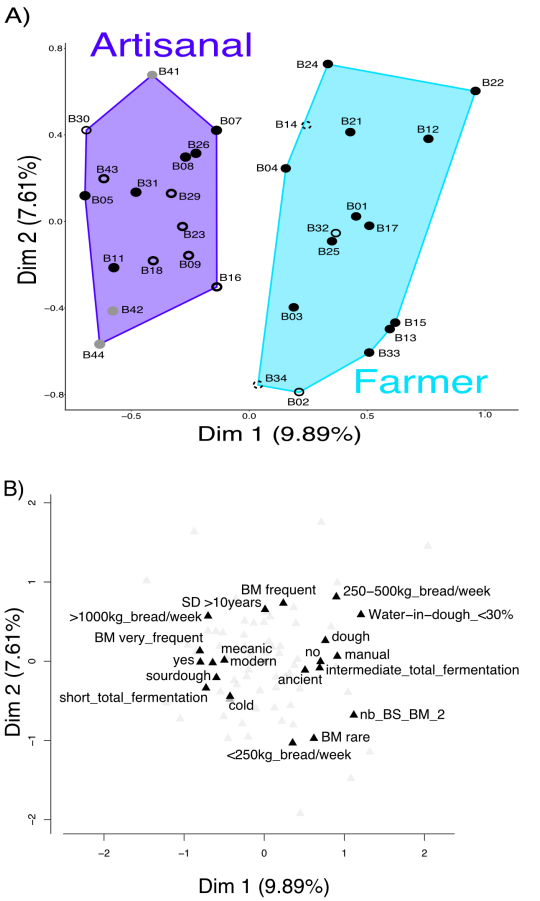

A)
Artisanal
Farmer
Dim 1 (9.89%)
Dim 2 (7.61%)
B)
BM frequent
250–500kg_bread/week
SD >10years
>1000kg_bread/week
Water–in–dough_<30%
BM very_frequent
dough
mecanic
modern
yes
no
manual
intermediate_total_fermentation
sourdough
ancient
short_total_fermentation
cold
nb_BS_BM_2
BM rare
<250kg_bread/week
Dim 1 (9.89%)
Dim 2 (7.61%)
C)
Artisanal
Farmer
Chief origin
dough
sourdough/dough
sourdough
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

Figure 3

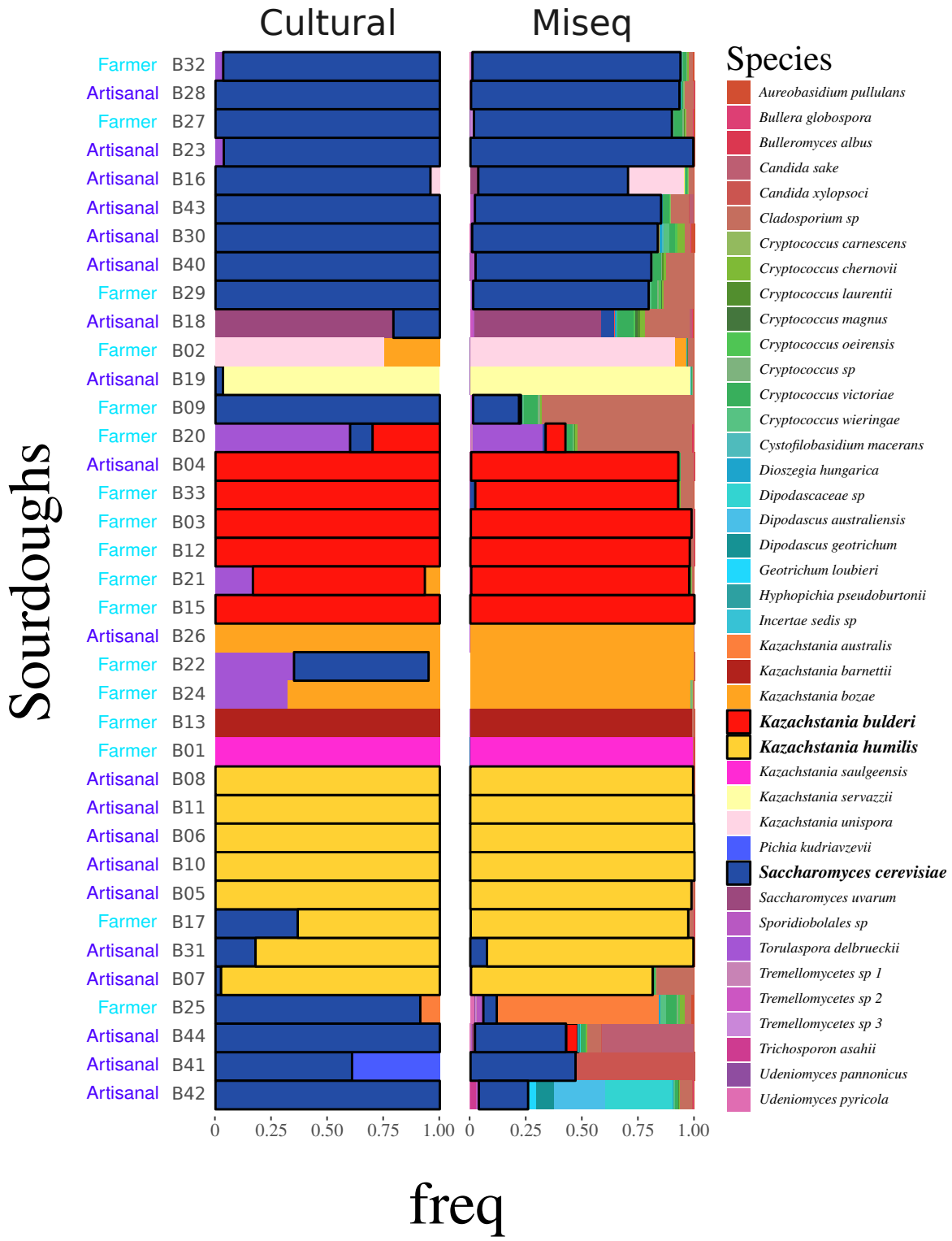

Cultural
Miseq
Species
Sourdoughs
Farmer B32
Artisanal B28
Farmer B27
Artisanal B23
Artisanal B16
Artisanal B43
Artisanal B30
Artisanal B40
Farmer B29
Artisanal B18
Farmer B02
Artisanal B19
Farmer B09
Farmer B20
Artisanal B04
Farmer B33
Farmer B03
Farmer B12
Farmer B21
Farmer B15
Artisanal B26
Farmer 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
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

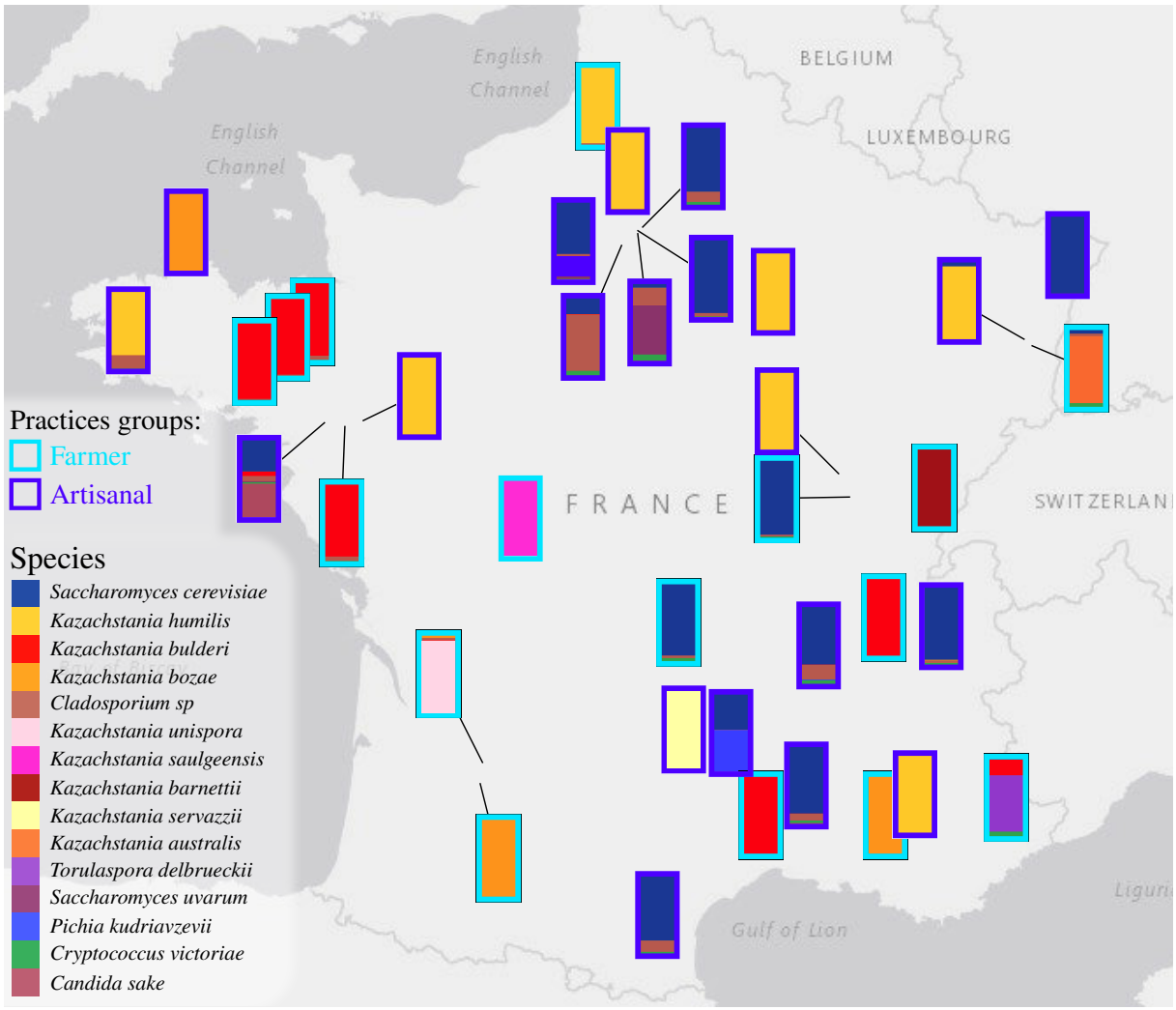

English Channel
English Channel
BELGIUM
LUXEMBOURG
FRANCE
SWITZERLAN
Gulf of Lion
Practices groups:
Farmer
Artisanal
Species
Saccharomyces cerevisiae
Kazachstania humilis
Kazachstania bulderi
Kazachstania bozae
Cladosporium sp
Kazachstania unispora
Kazachstania saulgeensis
Kazachstania barnettii
Kazachstania servazzii
Kazachstania australis
Torulaspora delbrueckii
Saccharomyces uvarum
Pichia kudriavzevii
Cryptococcus victoriae
Candida sake

Figure 5

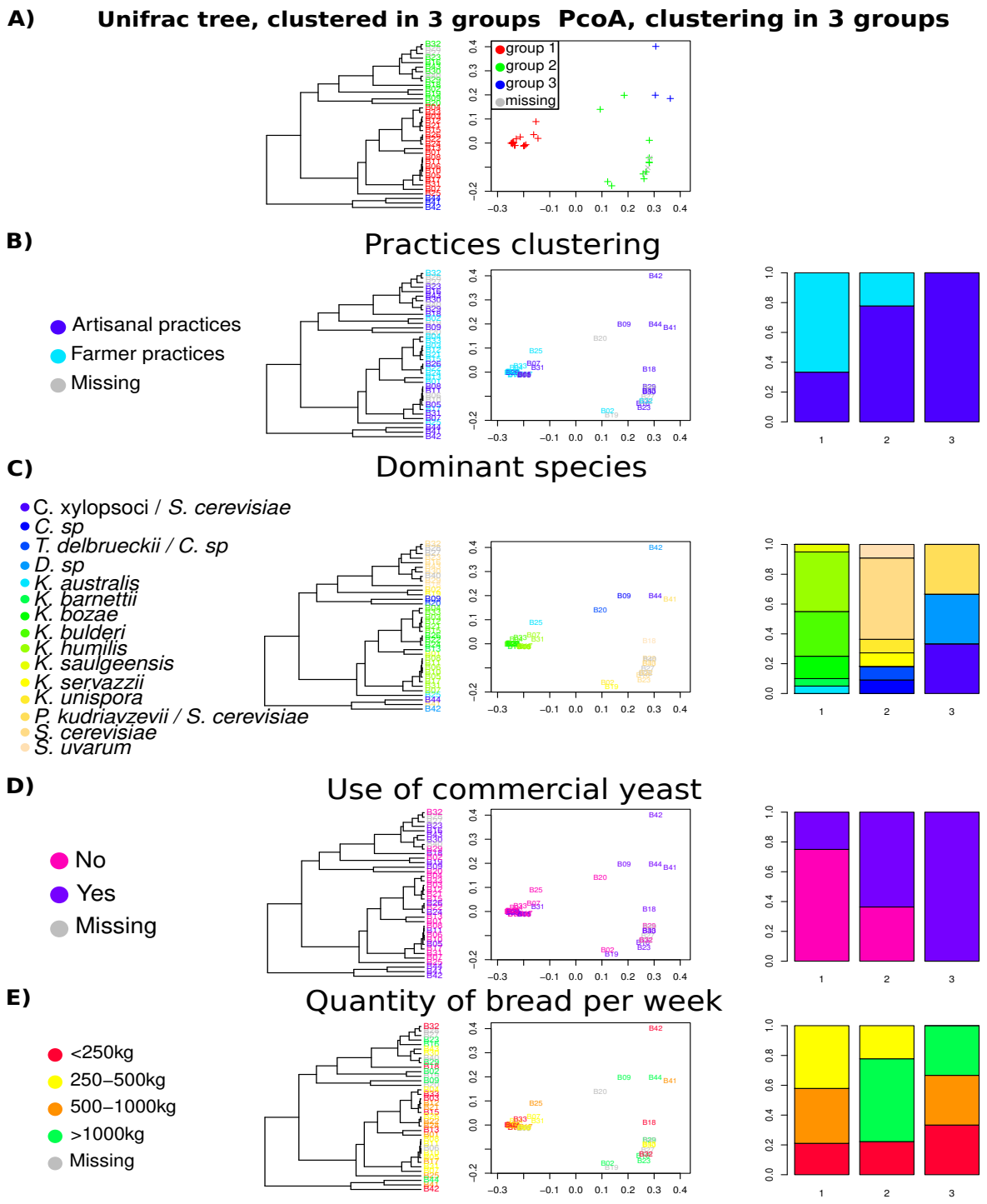

A)
Unifrac tree, clustered in 3 groups
PcoA, clustering in 3 groups
group 1
group 2
group 3
missing
B)
Practices clustering
Artisanal practices
Farmer practices
Missing
C)
Dominant species
C. xylopsoci / S. cerevisiae
C. sp
T. delbrueckii / C. sp
D. sp
K. australis
K. barnettii
K. bozae
K. bulderi
K. 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
250–500kg
500–1000kg
>1000kg
Missing

Figure 6

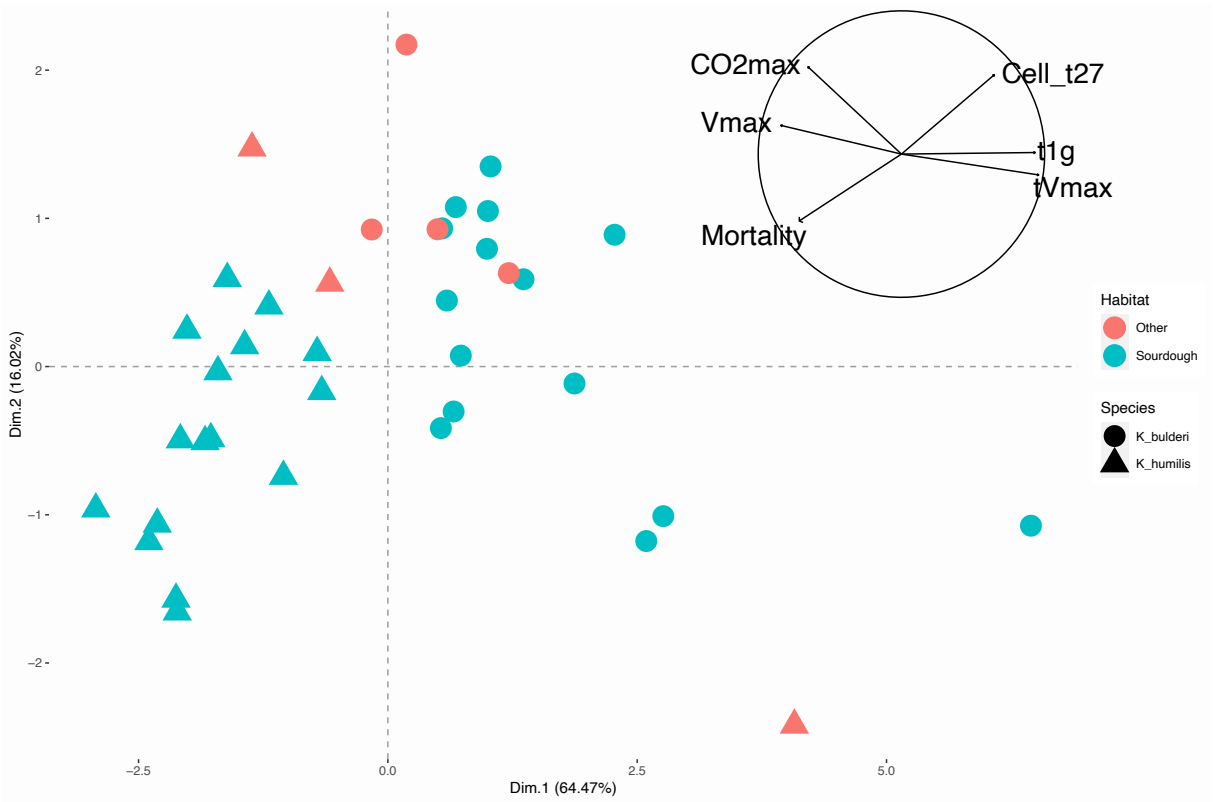
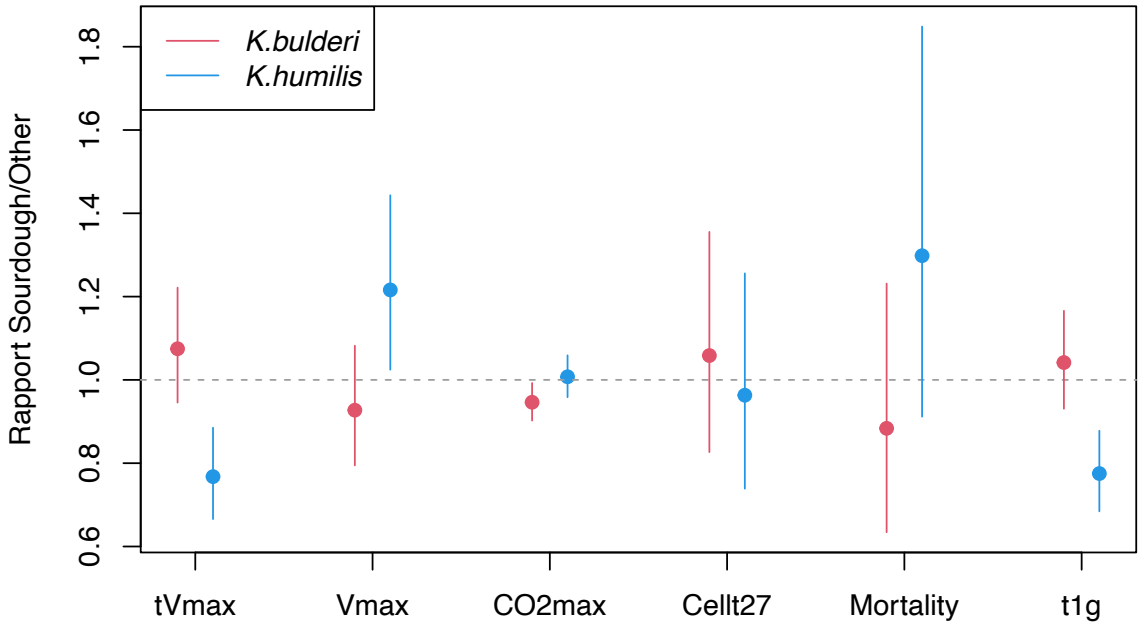

CO2max
Cell_t27
Vmax
t1g
tVmax
Mortality
Habitat
Other
Sourdough
Species
K_bulderi
K_humilis
Dim.2 (16.02%)
Dim.1 (64.47%)
2
1
0
−1
−2
−2.5
0.0
2.5
5.0

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