

The revised manuscript has been greatly improved, but I still have some suggestions before the manuscript can be recommended, in particular about the clarity of the text and exactitude of the terms.

Thank you very much for all your suggestions and English corrections.

We have made all the request corrections and have done our best to answer your comments.

We have submitted a new version of our manuscript on BioRxiv :

<https://www.biorxiv.org/content/10.1101/679472v4>

We have also submitted a tracked change document on the PCI, with our modifications in red.

As suggested, we have tested the effect of bread making practice on the level of alpha-diversity in sourdough. We found no significant effect of the bread making practice group on alpha diversity and added this result L327 “ The bread making practice group (artisanal-like/farmer-like) did not influence significantly the level of fungal α -diversity in sourdough (Wilcoxon rank exact test, $W_{shannon} = 156$, p -value = 0.16, $W_{simpson} = 165$, p -value = 0.078).

-L40: unclear: do you mean diversity within species or species diversity? Both are important and threatened and could be highlighted

We changed the sentence by “allow the conservation of a wider species and genetic diversity.”

-L45: why surprisingly? explain or delete

done

-L47: put "of the closely related *Kazachstania* genus » between brackets, otherwise it means that *Saccharomyces* is included in the *Kazachstania* genus

We replaced it by “while other yeast species belonging to the *Kazachstania* genus were dominant in 54% of sourdoughs »

-L48 : unclear « the distribution of yeast species » : distribution across what ? among what, or do you mean composition ?

We have added “across sourdoughs.”

-L49 : differences between what and what ? Do you mean « the most striking bread-making practice effect « ?

Thanks for this proposal. We replaced our text by this one.

-L51 : syntax issue in « Phenotyping of these two species in laboratory sourdough mimicking media revealed » and sounds partly redundant with the rest of the sentence

We have deleted “Phenotyping of these two species in laboratory sourdough mimicking media revealed phenotypic divergence...” and replaced it by “Phenotypic divergence between sourdough and non-sourdough strains were found for *K. humilis* but not for *K. bulderi*.”

-L53, L521 : no plural at practices as it qualifies diversity

corrected

-L54 : this is not a problem of taxonomy (classification) but of biological diversity (eg species diversity)

We replaced taxonomy by “species” diversity

-L66 : « again » does not seem the right term here, better to just delete (or explain when was the first episode of conservation this implies ?)

We have deleted “again”

-L67 : distribution across what ? among what ? or just « on species composition » ? (or diversity ? this is very different and it’s unclear here what you mean)

We have added “across sourdoughs”

-L70 : « The origin ... originated » : redundant

We have deleted “the origin”

-L71 : the reader miss here a date for plant domestication and what was the first domesticated plant

We have added a sentence on the cereal that were used to make bread during the Neolithic. “Investigation of the morphology of plant remains which were incorporated in Neolithic bread identified wheat, barley, millet, linseed [12].” But we prefer not to give a date of plant domestication. Records of human processing cereal date back to 13000 years ago in Israel and 14 000-30 000 years ago in Australia. There are some evidences that Australian were developing agriculture at that time. But it is unclear whether their activities resulted in the selection of domesticated form of cereals. Therefore, we find it difficult to give a date of domestication.

-L72 : not sure about the term « combined » here... « linked » ? « associated » ?

We changed it by “associated”

-L76 : is « process » the right term here ? you mean a here a particular event not the way of doing it

We replaced it by “maintained over time”

-L82 « and have improved » sounds too much active and anthropomorphic, replace by « which has improved » (domestication has improved fermentation, it's not the yeasts that have decided to improve their fermentation)

We replaced it

-L83 : why « however » ? delete ?

deleted

-L90 : *Pichia* in italics

corrected

-L91 : distribution across what ? among what ? or just «presence in sourdough » ?

We replaced it by “presence in sourdough”

-L94 : « or » instead of a comma before house

We deleted house characteristics, as we found it “too vague”

-L96 : professional bakers plural

corrected

-L99-100 : farmers-bakers or farmer-bakers ? choose one and keep it consistent

We chose farmer-bakers and corrected all

-L139 : with instead of according to

corrected

-L143, L191, L202, L415 : no capital within sentences even for defining acronyms

corrected

-L147 : no hyphen in 33-yeast

corrected

-L159, L182 : no s at OTU as it qualifies another, following word

corrected

-L162 : was assessed

corrected

-L181 : fungal communities, or just delete « To analyse fungal community » that is redundant with the following

deleted

-L183 : trees

corrected

-L185 : why did you test different roots and how did you assess which one to choose ?

We chose the roots among the OTUs that were affiliated to the most distant taxa, and thus belonging to basidiomycota. As tree architecture did not change with roots, we could choose any root. **The results presented in the main text were obtained using the phylogenetic tree rooted on the OTU identified as *Sporidiobolales* species.** We have added this information in the text now

-L186 : « It did » : who ?

“It “ was replaced by “The tree”

-L187, L193 : a tree does not « classify » taxa, it represents evolutionary relationships ; classifying is a human decision.. « misassigned to » ?

We replaced “some *Ascomycota* were misclassified” by “some *Ascomycota* were located among the basidiomycota”

-L188 : « in » instead of « according » and delete « were »

corrected

-L196 : define Chao1

We have added the sentence: “Chao1 was used as an indicator of species richness corrected by the number of OTUs present in the community but not observed”

-L197 : index as it qualifies another, following word

corrected

-L203 : I doubt you really performed the analysis on the bakers themselves ?

We replaced “we performed the analysis on the 30 bakers” by “We included in the analysis sourdough fungal communities of the 30 bakers”

-L204, L206, L208 : practice as it qualifies another, following word ; define FDR and add a reference

FDR was defined few lines below, we defined it here now.

-Overall, the passive form is much too used, it renders the text cumbersome to read. The active form should be preferred, it's lighter and more dynamic.

We modified the text as suggested and made the explanation clearer.

-L208 : I guess you do not expect a single link between all these variables so plural would be better

corrected

-L220, L303, L361, L450 : delete « finally », it's not likely you only expected to arrive there

Thank you

-L222 : proxies

corrected

-L226 : fermentation latency phase (no hyphen there) ; define tlg

corrected

-L227 : cut the sentence

The sentence was cut.

-L235, L399 and elsewhere : all numbers below ten should be written in letters, except in equations

corrected

-L249 : R package (no hyphen there)

corrected

-L250 : move the bracket at the end of the sentence, it will be easier to understand

corrected

-L261 : unneeded repeat of « bread making process »

deleted

-L266, L275 and elsewhere : keep the past tense

corrected

-L268 : what does « ancient wheat populations » mean ? harvested centuries ago ?
Otherwise it would rather be « ancient varieties » ? or what ?

Variety often refers to genetically homogeneous population, while ancient varieties are often grown as a mix of genetically diverse populations. That's why we called it population. To avoid confusion, we replaced wheat populations by landraces.

-L270 : why using « baker's yeast » ? starter ? *S. cerevisiae* ? other yeasts ?

They use *S. cerevisiae* as starter. We therefore changed the text to "using *S. cerevisiae* starters"

-L282 : delete « see M&M section » but explain when this is needed

We deleted « see M&M section » and added "when the ITS alone was not able to discriminate between closely related species"

-L288 : coma instead of space at 194 557

corrected

-L289 : « were identified as non-yeast » : really ? by molecular markers ? « yeast » is not a taxonomic status, it is a growth form opposed to « filamentous », there are also basidiomycetous yeasts ; same issues at other places in the manuscript, e.g. L292, L297 (and I don't know what yeast-like means)

We completely agree. Indeed, by saying non-yeast, we wanted to speak about OTUs that were assigned to either filamentous fungi or plant. To avoid confusion, we deleted the term non-yeast. We replaced the sentence by: "Among all OTUs, 10 were assigned to the order *Triticodae* (especially to the species *Triticum aestivum*), 50 were assigned to a filamentous fungi genus including plant pathogen species such as *Alternaria*, *Aspergillus*, *Fusarium*, or *Gibberella*, while 4 OTUs remained unidentified."

-L306 : unclear what you mean exactly by « distribution of fungal species diversity » ? what does « distribution of » add ? across what ?

We replaced by "the distribution of fungal species across sourdoughs"

-L308, L309 and elsewhere : explicit the genus when you did not mention the species yet, it is unclear what you're referring to

We replaced by *Cladosporium* genus

-L310 : what do you mean by « distribution pattern » ? what is the question ? I have a hard time with this term, I don't understand what you mean when using it ? It seems like just a vague term lacking clear questions.

We replaced the term pattern by “species composition of fungal community”

-L313 : lower relative abundances : I guess they don't all have the same abundance value ?

corrected

-L316 : differences in terms of what ? species presence/absence ? species diversity ? abundances ?

We replaced it by “ β -diversity”

-L320 : keep the same difit numbers

done

-L320 : shown

corrected

-Figure 1 : why all this white space ?

We made a new figure

-Figure 2B : the text is still too small to be readable

-The figure 4 could probably be improved

We already improved figure 4 several times to fit reviewer's request. We have kept the species color code constant for all figures. We'd rather prefer not to change it, also to avoid mistakes.

-Figure 5 : labels are not readable on trees and PCAs ; give whole species names without abbreviation (also Figure 7) ; « Practices clustering » : practice clustering or just practices ? « unknown » instead of « missing » or explicit what is missing

We replaced “practices clustering” by “practice clustering”. We now gave the whole species name in Figure 7. In Figure 5, we could not change the species name because it took too much space but we explained species name abbreviation in the legend.

-L323 : again unclear what you mean by « distribution » ; yeastS (or just fungi ? « yeast » is not a very precise term)

We replaced it by “occurrence of yeast species in sourdoughs”

-L325 : again unclear what is a « yeast species » ? do you have a taxonomic meaning ? or just growing as unicellular ? or able of fermentation ? you should define at the very beginning of the manuscript what you mean by yeasts ?

We now defined yeasts in the introduction: “Yeasts, whether ascomycetes or basidiomycetes, are generally characterized as fungi, that asexually reproduce by budding or fission, which results in growth that is comprised mainly of single cells. Their sexual states are not enclosed in fruiting body.”

-L341 : delete « indeed »

deleted

-L363 : replace « evidenced » by « reported » or « observed »

Replaced by observed

-L365 : be consistent all across the ms in « bread-making » with or without a hyphen

We chose without hyphen

-L368 : fewer instead of less

corrected

-L370 : keep the past tense

corrected

-L381 Group 1 encompassed

corrected

-L384 : harbored sorgoughs with

corrected

-L389, 390 and elsewhere : for each test, do not just give the P value, but also the degree of freedom, the statistic value and all information needed for repeatability and check

Exact fisher tests are done on contingency tables, there is no ddl. When we made test based on F-statistics distribution, we gave the ddl (ie line 437) .

-L429 : farmers-likers ?

corrected

-P18-19 : all these results are a bit cumbersome to read and it is difficult to extract the main message.. would it be possible to summarize also on a figure the main features of the different groups ?

We summarized this part by “In conclusion, no clear evidence was found of the impact of bread making practices or of the dominant yeast species on the metabolic composition of sourdough. On the other hand, our results showed metabolic differences between sourdoughs having one or two co-dominant yeast species.”

The PCA that we have presented as supplementary information give a general picture of these results.

-L446 : avoid this abbreviation

We could not find which abbreviation had to be changed

-L447 : fungal and bacterial would be more consistent

corrected

-L479 : comparisonS or A comparison

corrected

-L491 : unclear what « these findings » refers to ? and « the diversity found in France » : do you mean in the present study ? This sentence is highly unclear

We deleted the first part that was cumbersome. The sentence became “All the yeast species detected at a relative abundance over 1% in this international collection of sourdoughs were detected in French sourdoughs except the species *Wickerhamomyces anomalus*, *Pichia membranifaciens*, *Naumovozyma castellii* and *Saccharomyces bayanus*.”

-L493 : delete the comas before *Wickerhamomyces* and after *castellii*

corrected

-L495 : replace the coma after *australis* by « and »

corrected

-L496 : do no compare a species and a genus

We replaced the sentence by “Beyond the genus of the baker yeast species, *S. cerevisiae*, the most represented yeast genus in French sourdough was *Kazachstania*.”

-L498 : if you list them all, « include » is not the right word

corrected

-L502 : avoid the unclear « it » by fusing the two sentences

corrected

-L504-5 : unclear why this is an issue to study genetic diversity

We changed our sentence. Here we wanted to say that “its taxonomic characterization may have been hampered by the fact that it probably originated from hybridization between unknown yeast species”

-L506 : delete « the » unless you mean a specific sourdough ? but which one ?

corrected

-L508 : a bit strange to cite only cichlids here ? cite other cases, and at least a few of domesticated organisms, including fungi

Yes, we agree. We have added the example of *Penicillium*.

-L511: replace the comma after bulderi by « and »

corrected

-L513 : this is a bit short and unclear

We have added “and the genetic changes that would have been selected during potential domestication »

-L522 : you showed that it's higher but not that it increased, you don't know the ancestral state

corrected

-L531 : similarly, delete « of selection » : you would have only be able to detect different levels not whether this would have been due to selection

corrected

-L536 : twice « during the week » while neither seems important

deleted

-L544 : as already said, silage, not ensilage

corrected

-L561 : not sure that « diverge among themselves » is correct or mean something ? diverged one from each other ?

We replaced by “from each other”

-L565 : and instead of et

corrected

-L576 : « important » has not the same meaning as in French, it is not quantitative but qualitative, and bottlenecks

We replaced “important bottleneck” by “strong” bottlenecks

-L577 : italics at species name (also L580, 584) and add soft-cheese before *P. camemberti* to avoid confusion with the adjective blue cheese placed before

corrected

-L578 : diversity erosion in the domesticated organisms used for fermented product making ; and maybe you can be more specific by citing the studies having revealed degeneration in domesticated fungi (eg loss of sex ability) ; this would help justifying your conclusion where you say that we need to preserve microbial diversity in food. Some additional examples of the benefit of having microbial diversity in food would be good.

We now spoke about fertility depression in domesticated fungus, as followed: “The risk of genetic diversity erosion can also be accentuated by fertility depression among fungal domesticated strains [86]. “

-L580 : AOP is PDO in English and should be defined (protected designation of origin) ; maybe cite the *P. camemberti* study (Ropars 2020) for the issue associated with PDO but Dumas et al 2020 showed in contrast that PDO had protected diversity in the Roquefort *P. roqueforti* population.

Thank you, we corrected it and cited Ropars et al. 2020

-L584 : had ; different from

corrected

-L586 : a reference is missing, but if you mean the Ben Wolfe’s study, this study is actually flawed in its conclusion as discussed in the Ropars 2020 paper : the strain used as collected in a cellar where cheese is cultivated and it belongs to a domesticated clade, so this was just not a wild-type strain

We meant Ben Wolfe’s study. We changed the sentence to “A recent study suggested that domesticated *Penicillium* strains can evolved phenotypically in a few weeks [87]”

