

Dear Christelle Fraïsse,

Thank you very much for taking the time to consider our manuscript for recommendation and for the constructive review process. We have revised the manuscript according to the reviewers' comments and highlighted in yellow the changes in the manuscript. We outline below (in blue) how we addressed all comments and we think this improved the manuscript, which we are grateful for.

Best regards,

**Bastien Bennetot, in behalf of all authors**

### **Decision**

by *Christelle Fraïsse*, 21 Oct 2022 17:36  
Manuscript:  
<https://www.biorxiv.org/content/10.1101/2022.05.17.492043v3>

### **Minor revisions needed**

Thank you for your patience. Two reviewers have provided constructive and thorough comments on your preprint. Both agree that the study is of high quality, and after reading the manuscript, I agree with their statement. I especially appreciated that domestication was tackled from a phenotypic and genomic perspective - this represents an impressive amount of work. I am convinced that this work will interest the evolutionary biology community.

>> We thank you for these positive comments.

The reviewers made a number of helpful, often concordant, suggestions for improvements that must be addressed in a revision. In particular:

1) They criticize the interpretation of the data for the Cheese\_2 population as revealing a more advanced stage of domestication. Evidence for human-mediated selection to be more substantial in that cheese population than others should be demonstrated more clearly.

>> We have modified the text to show more clearly that the domestication syndrome is stronger in the cheese\_2 population, with a stronger genetic bottleneck, denser and fluffier colonies and a better

capacity of inhibiting cheese spoiler fungi, corresponding to phenotypes more typical of cheese fungi. We have nevertheless tempered our claim and have added discussions about alternative hypotheses.

Moreover, Reviewer 1 suggests considering other forces than selection in the evolution of the cheese populations, including migration and genetic drift.

>>We have made changes throughout the text to suggest that migration and genetic drift may also have shaped the evolutionary history of cheese populations of *G. candidum*. Note however that we do not have much evidence for migration.

2) Another issue raised by the two reviewers is the under-sampling of the wild strains. This bias could have consequences for interpreting the data (in particular, if the wild strains were sampled in a single geographic area), so it should be handled with caution. Reviewer 2 noted that an outcome that could follow is the misinterpretation of the cheese clade as being derived from the wild clade (see their specific comments on that point). Determining if cheese-making practices (Reviewer 1) could explain the differentiation between the three cheese clades may help understand whether this genetic structure within cheese strains has pre-existed domestication.

>> We agree with reviewers that wild strains are under-sampled. However, the 10 wild strains available captured a substantial diversity both in terms of substrates (e.g. soil, flower and polyurethane), and of geographic origins (i.e., Thailand, UK, French Guiana, Brazil, Egypt, Senegal, South Africa, Belgium, Spain and Sweden). We have highlighted this point in the manuscript. We have nevertheless softened our conclusions to take into account the undersampling of wild strains throughout the manuscript. If cheese-making practices have shaped the population structure of *Geotrichum candidum*, we expect an association between cheese type and population. In fact, all soft ripened goat cheeses clustered in a single population (Cheese\_1). However, the two other cheese populations contained strains isolated from all types of cheeses (fresh, hard and soft mold-ripened). We have added a column in the Table S1 for types of cheeses and we discuss this point. We have however clarified that we cannot exclude that the genetic subdivision pre-dated domestication, as we have few wild strains.

3) Both reviewers note the difficulty of interpreting  $F_{ST}$  when measuring the extent of population divergence. Moreover, Reviewer 2 questions the method used to detect candidate regions under adaptation. Instead of considering the 5% most extreme values of  $D_{xy}$  and  $\Pi$ , which is expected under neutrality, you could only consider those regions that pop up as extremes in both  $D_{xy}$  and  $\Pi$  scans as candidates. Otherwise, I recommend running a method that scans for selective sweeps.

>> We now consider regions that pop up as extremes in both  $D_{xy}$  and  $\pi$  scans as candidates. We also ran SweeD for identifying selective sweeps, and only considered those that matched regions already identified as extremes in  $D_{xy}$  and  $\pi$  scans. We have modified the text accordingly.

4) I agree with Reviewer 1 that the comparison with *Penicillium* should be better justified and moved into a single paragraph in the Discussion section. A more in-depth discussion of the convergence process would help clarify the importance of comparing with *Penicillium*.

>> We agree and have modified the text accordingly, adding a paragraph in the Discussion on the comparison with domesticated *Penicillium* fungi.

5) Reviewer 2 suggests running a demographic analysis to test for an absence of a substantial bottleneck during the domestication of *G. candidum*. I think there is no need to run such a type of analysis; however, Tajima's  $D$  could be calculated to clarify this claim.

>> We have computed Tajima's  $D$  and added values in our manuscript.

I am looking forward to receiving your revised preprint.

With best regards,

Christelle Fraïsse.

## Reviews

*Reviewed by Delphine Sicard, 19 Oct 2022 07:47*

This manuscript reported the analysis of *Geotrichum candidum* domestication in cheese making. This MS expands the studies on the domestication of fungi in fermented products. It presents a set of results on the genomic diversity of the fungus and its phenotypic diversity. It shows that three groups of cheese populations have diverged genetically and phenotypically from populations with wild and mixed-origin and revealed some interesting features on the evolutionary trajectories of the cheese populations. I found that the work done was adequate with the scope of the manuscript. The analysis of the interactions between *Geotrichum candidum* and its competitors is very original in a study on domestication. My comments are just made to improve the manuscript.

>> We thank the referee for this positive assessment.

Majors :

I suggest to use the term domestication with more care. Please define domestication and discuss the results with respect to selection by cheese making practices and to migration via the use of industrial strains.

>> The term domestication was defined at the beginning of the introduction L49-52: "Domestication, the change in the genetic and phenotypic make-up of populations under human artificial selection, is an excellent model for studying adaptation processes, as it involves recent adaptation events under strong selection on known traits, rapid diversification." We have added "and reduced gene flow between wild and domesticated populations" at the end of this sentence (L52-53).

Similarly, the paper refers several times to phenotypic convergence but the processes underlying this convergence should be explained. In other worlds, it would be interesting to propose the different scenarios that could explain these convergences.

>> We have clarified that we only studied phenotypic convergence and we do not have any clues yet on the underlying processes. We have added a sentence in the conclusion stating that future studies should tackle this question of the underlying genomic processes.

We are left with not knowing if the *Geotrichum candidum* cheese populations have adapted to the high lipid content of cheese. The lack of lipolysis signature of domestication raised questions. The maladaptation of cheese populations to cheese agar medium, and to salt is intriguing and would deserve additional comments. Other hypothesis than evolutionary constraints could be suggested.

>> Indeed, the lack of signature of domestication for lipolysis, radial growth on cheese and salt is intriguing. We did discuss other hypotheses such as a lack of selection for faster growth if it leads to too degraded products. We have also added the hypotheses that selection may not be efficient enough for changing multiple traits at the same time and/or that migration counteracted selection.

More generally, the involvement of evolutionary forces other than selection in the evolution of the cheese populations should be further discussed. I would expect migration, especially those associated with the sale of starter strains. I would also expect drift in environments that are often disinfected.

>> We considered genetic drift as a driver of evolution in cheese populations when discussing bottlenecks; we have clarified this point in the manuscript. Our analyses showed no evidence of migration or gene flow (beyond admixed strains, that we discuss), as cheese strains grouped together separately from wild strains and we identified three genetically and phenotypically different populations within the cheese clade. None of

the F3 tests between pairs of populations was significant, supporting the lack of gene flow between populations. We have nevertheless highlighted that commercial starter strains are found in different cheese types, which suggest migration. However, as this does not result in gene flow, this has little impact on the “evolution of the populations”.

In terms of structure, there are several elements of discussion in the results, especially the comparison with *Penicillium*. I would move them into the discussion in order to reduce and clarify results and enrich the discussion.

>> We thank you for this suggestion. We removed all elements of discussion from the results and moved them to the discussion section.

Title - I would not call a group of genetically closely related fungal strains a variety. To me, a variety refers to a group of genotypes (most often a single genotype) that has been consciously selected by human.

>> We respectfully disagree here. While we understand that it sometimes refers to groups selected for different traits, it can also simply refer to groups with different phenotypes. It is even used in wild species as a synonym for subspecies. Moreover, in the case of *G. candidum*, we do suggest that there may be selection for different traits, even if we do not have definitive evidence. We have clarified what we meant by varieties and tempered our statements.

Minor

L28-29: “The genetic diversity ...was high”: high compared to what ?

>> We have clarified the sentence.

L29: I would not state in such determine way that the data indicates a lack of strong bottleneck because of the sampling biases. There are many more cheese strains than strains coming from elsewhere.

>> We have softened the claim, highlighting the sampling bias. However, we note that we still observe higher genetic diversity within the cheese clade than in other domesticated cheese fungi. We have also clarified that the wild strains come from various countries in different continents and substrates.

L32 “attractive” for who ?

>> We have modified the sentence: “a prominent production of typical cheese volatiles” L32.

L34 what do you mean by “a more advanced state of domestication” ? I don't think the data allows the quantification of the response to selection . Moreover, the decrease in genetic diversity is not a signature of

domestication. Several studies have now shown that domestication may also lead on the contrary to a diversification, at least for some traits or genetic clusters.

>> We have clarified what we meant and softened our claim. We have tried to clarify that there seems to be a stepwise process of domestication, as documented in multiple crops, with a gradual increase in the domestication syndrome, i.e., a gradual differentiation in phenotypic traits from the wild population. Although we cannot be completely certain, we think this is worth highlighting as a possibility, and we now discuss alternative hypotheses, in particular that differentiation predates domestication. Nevertheless, we explain more clearly that the Cheese\_2 population shows more differentiated traits from wild populations, with multiple traits similar to other cheese fungi and beneficial for cheese making..

L74-75: *S. cerevisiae* has been domesticated to make a large number of fermented products beyond beer, wine and bread. Please complete and cite either a complete set of recent reviews on *Saccharomyces cerevisiae* domestication or a complete set of major research papers in the field. I would go for research papers and add at least Barbosa et al. 2018 (sake), Bigey et al. current biology, 2021 (bread), Gonçalves et al. current biology, 2016 (beer, wine), Ludlow et al. current biology, 2016 (coffee, cacao)

>> We thank the reviewer for pointing out these missing papers in our citation list. We have added major research papers in the field L75.

L105-if you really want to use it, define degeneration and cite a reference. Personally, I find it more informative to speak about accumulation of deleterious mutation if this is what is meant.

>> We thank you for the suggestion; however, we used the term degeneration to refer to the phenotypes while deleterious mutations refer to the genotypes, although, of course, the latter is the cause of the former. Our previous papers demonstrated degeneration in fertility for example, but we have not formally demonstrated an “accumulation of deleterious mutations” in genomes and we are therefore afraid that such a formulation may be misleading. We have modified the sentence in the text L98-102 to make it clearer (note that what we mean by degeneration is explained in the sentence): “Bottlenecks (leading to genetic drift) and degeneration have also been documented in domesticated fungi, with reduced fertility and genetic diversity in the cheese fungi *P. roqueforti* and *P. camemberti*<sup>33,35</sup>, likely due to an accumulation of deleterious mutations because of drift.” We have also clarified the term elsewhere in the manuscript.

L125....--it is a repeat of the abstract. If you wish to add a summary of the results at the end of the introduction, it should be more concise and differ from the abstract.

>> We reformulated and shortened the summary of the results in the introduction. Note however that the end of the introduction should introduce all aspects of the work to guide the reader through the material & methods and the result section, so all aspects should be present for optimal clarity for readers. That the abstract is redundant with some parts of the manuscript is not an issue, in our opinion, as the abstract is precisely a summary of the manuscript.

L186- clarify “cheese type distribution”

>> Cheese type refers to soft (e.g., Camembert), soft natural rind (e.g., fresh goat cheese), pressed cooked (e.g., Comté), pressed uncooked (e.g., Tomme) and blue cheeses. We have clarified this term in the text L187-189.

L189-103- move the comparison with *Penicillium* to the discussion. In addition,  $F_{ST}$  are used to measure the extent of divergence among populations of the same species relative to the net genetic diversity within the species. To compare species, absolute measures of divergence between populations should be used in preference to relative measures such as  $F_{ST}$ . (Charlesworth’s paper and others, DOI: 10.1093/oxfordjournals.molbev.a025953)

>> We thank you for the suggestion. We have moved the comparison with *Penicillium* fungi to the discussion and now compare  $D_{xy}$  values instead of  $F_{ST}$  values.

L232-233: avoid suggestion in the results, move to the discussion

>> We have moved this suggestion to the discussion section.

L243: I can’t see the yellow Figure 1B.a., and therefore the position of commercial strains on the tree. Please, complete the Figure 1 legend as well

>> We are sorry, we forgot to correct the text L243 after a change in the figure. Commercial strains are pinpointed with a “\$” symbol. The legend of figure 1 is correct.

L247- 1,200 SNPs: how did you choose this threshold ?

>> We did not choose a priori a threshold of 1,200 SNPs, but instead considered as clonemates the strains clustering together in the ML tree without branch length. We then checked the level of polymorphism between these clonemates, and found that it was less than 1,200 SNPs in all cases. We have clarified this point in the text L256-263.

L252: move to the discussion

>> We have moved this part to the discussion.

L242-254 or elsewhere: I was not able to find any information on ploidy. Are all the strains haploid ? please, add the information.

>> All strains considered in this study are haploid. We have added this information L150-151.

L350-avoid example on Penicillium in the results, move to the discussion

>> We have moved this example to the discussion.

L353-356- what does “harsh conditions” mean ? this part should go to the discussion. Furthermore, the relaxed selection hypothesis applies to industrial strains but does it really apply to non-industrial cheese strains?

>> We now give a definition for “harsh conditions” in the text L391. We think that this sentence is important here in the Result part because it explains one of our tested hypotheses. The hypothesis also applies to non-industrial strains as they also grow in the nutrient-rich cheese matrix.

L357-I would have expected “local adaptation”, i.e. a higher growth on cheese media of cheese strains compared to wild strains. I would discuss this result further in the discussion

>> This is indeed the hypothesis we aimed at testing, and we now discuss this surprising result more completely in the discussion section.

L369- move the convergence analysis to the discussion session.

>> We have moved all the discussion on convergence in the discussion section.

L411: delete “known to be key compounds in fermented beverages such as wine and beer”

>> We have deleted this part of the sentence.

L452-456: discussion

>> We have moved this part to the discussion.

L485 / L498: Why do you compare the Geotrichum candidum cheese populations diversity with the ones of Penicillium sp. ? did you have any prediction based on the use of these species for cheese making ? What does the Penicillium and Geotrichum diversity comparison brings ? I would made a single paragraph comparing both genera rather than speak about it all over the discussion. This would allow to better show the hypothesis you have to explain their evolutionary trajectories.

>> We thank the reviewer for this suggestion. We have remodeled the discussion part and created a “convergence” section L673-707.



L503-506: the secondary domestication hypothesis can be discussed in a more general context of domestication. There is the example of the cachaça *S. cerevisiae* populations (Barbosa et al 2018) but other examples could also be added.

>> We have added this example and we discuss this point in more detail L638-640.

Does cheese making practices could explain the differentiation in three groups of the cheese strains ?

>> See our answer above: this may be the case for the Cheese\_1 cluster but it does not seem to be the case for the two other clusters. We instead suggest that the differentiation of the Cheese\_2 cluster results from a selection for *P. camemberti*-like features.

L533-L534: What do you mean by “ first step of domestication” ? wby “a more advanced state of domestication” ? There is no dating here, no analysis of evolutionary dynamics. I think you can't tell where the cheese populations are on the adaptive peak. You may also have several adaptive peaks.

>>We have clarified that we meant that the three cheese clusters display different levels of phenotypic differentiation from the wild clade, with a stronger phenotypic differentiation and a stronger bottleneck in the cheese\_2 clade. We agree the term “first step” could be misleading and we have reformulated the text and tempered our claims.

L555- Why is “convergence” an important question in evolution ? a deeper discussion on the interest of studying convergence would be interesting. Explaining the process behind phenotypic convergence (standing genetic variation, de novo mutation, migration/gene flow, horizontal gene transfert) would help to understand your approach.

>>We agree this would be fascinating to know but we have for now too little knowledge on the genetic basis of the studied traits to discuss this without too much speculation. We only meant phenotypic convergence, which we have tried to clarify. We have nevertheless added a sentence stating that it will be interesting to investigate these questions in future studies L721-725.

Figures legend-please describe in more details your figures in their legend, so they can be understood as themself.

>> We describe our figures in more details in the legends.

[Download the review](#)

Reviewed by anonymous reviewer, 17 Oct 2022 12:29

Bennetot *et al.* studied patterns of genome-wide diversity and population differentiation in the fungus *Geotrichum candidum* used for cheese-making. They analysed genomic data from 98 strains as well as phenotypic data from a subset of strains. They found clear population structure, clustering strains into three main clades, one of them composed by cheese making strains. Additional population structure was found within the cheese making clade, with three additional subclades, but also with signatures of admixture between them. Interestingly, they found higher genetic divergence between cheese making strains (in comparisons between subclades) than between wild strains, interpreted as reduced bottleneck during domestication. Cheese making strains showed phenotypic differences relative to wild strains, in traits like growth, colony morphology, volatiles production and proteolytic activity. Finally, the authors identified genomic structural variation between clades, including variation in gene content and transposable elements.

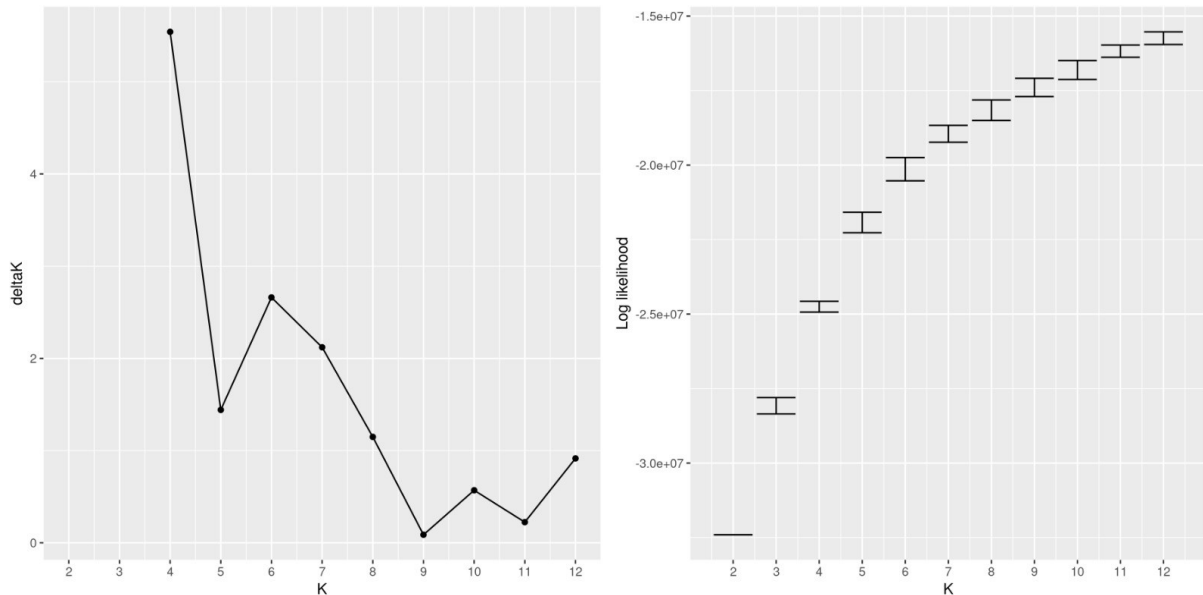
I think this is excellent work. The manuscript is interesting, well-structured and written. There are several very interesting, and even unexpected findings. I believe this work is relevant not only within the fungal community but also in a broader audience within the context of domestication. I learned a lot and enjoyed reading this manuscript. However, I also found that the analyses done in the study are often insufficient to support some of their claims. I was also confused with some of the discussion points, which I believe should be clarified. Please see below my comments.

>> We thank the referee for these nice words.

- The authors performed an admixture analysis to identify population structure. Although using  $K=5$  populations are consistent with monophyletic clades in the phylogeny, an adequate statistical/maximum likelihood test is not performed to define the potentially optimal value for  $K$ . They stated "At higher  $K$ , new populations inferred were either too small (two individuals) or not monophyletic". I argue these are not adequate criteria to select the number of structured populations. An isolated population does not need to be monophyletic in the phylogeny in particular in a system where there is also admixture as they show. Additionally, a population can be represented by a single sample if this one has substantial divergence with other subpopulations

>> We have added the DeltaK statistic L166 (Evanno *et al.* 2005, DOI: 10.1111/j.1365-294X.2005.02553.x) to identify the strongest population subdivision level. Note however that we think that this statistic should not be said to give "the optimal  $K$  value" for natural populations. The way this statistic is designed gives the strongest level of subdivision (the level for which the differential in likelihood is the highest when increasing  $K$ ); the likelihood nevertheless still increases after this  $K$  value, and finer genetic

subdivision often genuinely exists that can be biologically more relevant. On the contrary, a lower subdivision level may be biologically more interesting, as in our case. This is why we did not compute this statistics initially, this statistics can be misleading when not interpreted properly and actually rarely gives the biologically most relevant subdivision level.



The second order rate of change in the likelihood ( $\Delta K$ ) peaked at  $K=6$ . The additional population distinguished at  $K=6$  compared to  $K=5$  encompassed two strains that were not that differentiated from others in the splittree (MUCL 14462 and CBS 9194; FIGURE 1B). Because we could not run phenotypic tests on a cluster with only two strains, and we cannot interpret cheese type or geographical origin with only two strains, we chose to consider only the five largest populations in the manuscript. We have added some sentences in the results to clarify these points L166-172.

- L186-196: I found it difficult to follow the comparison in divergence between the different cases. The Value provided is  $F_{ST}$  (a relative measurement of diversity within Vs between populations). I do not think  $F_{ST}$  values can be compared directly since the number of variant sites is potentially different for each case. I suggest to report  $D_{xy}$  values instead, as done for the second example.

>> We thank the reviewer for this suggestion. We now report  $D_{xy}$  values instead of  $F_{ST}$  values and we have streamlined the description.

- L199: It is quite puzzling that genetic diversity between domesticated species is higher than in wild samples. There is lower diversity within the cheese clades, potentially explained by domestication/selection. However, if it is considered that domesticated strains must have an ancestral wild origin, it should be clarified how wild strains could have lower diversity. Domesticated strains should have lower or the same in the most extreme

scenario. Does not this suggest a clear under sampling of wild strain, restricted to a small selection of related strains?

>> We agree with the reviewer that the diversity in the wild clade is likely low because of the few wild strains available. However, we have clarified that the wild strains are not just a small group of closely related strains from a specific origin: they originated from different continents and various substrates. We now discuss this point in the Discussion part explicitly L528-532.

- Since rooting in the tree was done by using middle point divergence, I believe the relationship between "ancestral" and "derived" can not be concluded directly from the tree. Domesticated strains are assumed to be derived from wild strains, but in my impression, with the limited sampling of wild strains this is not necessary the case. "Domesticated" genotypes, and in particular the three mayor clades could be ancestral groups which were subsequently selected more recently by humans (as shown by reduced genetic diversity within clades). Authors need to clarify what is the evidence to believe the three subclades within the cheese-making clade are derived, as opposite to ancestral variation within wild (but undersampled) population.

>> we thank the reviewer for pointing this issue out. First, note that all clades analysed here are present-day populations and none is derived from the other, there are no ancestral populations in our sample (such as could be old collections or fossils). We have clarified that the traits observed in the wild strains available may not correspond to the ancestral traits of the domesticated populations.

Second, we have run additional analyses to strengthen our inference on the tree rooting. We relaunched IQTREE version 2.1.3 using non-reversible models and tested root positions using rootstrap supports (<http://www.iqtree.org/doc/Rootstrap>). We think that the rooting of the *Geotrichum* clade is robust and that the cheese clade is a sister group to the wild clade (but it is not nested within the wild clade as could be expected if we had identified the population from which the cheese populations were domesticated). The wild and cheese clade are sister clades, we cannot say that one derives from the other one, the split is symmetrical. We have clarified our interpretation on this root. Nevertheless, we agree that the split between the three cheese clades may be more ancient than the domestication events, if we missed some wild populations. We discuss this point in the discussion L555-563, L641-643.

- Additionally, connected to the previous point, since the ancestry of clades is not clear, it is not clear to me how it is inferred CNV ancestry. I mean, samples show an increase or decrease in coverage of repeats relative to a particular reference. This could be an expansion in repeats in the different strains relative to the reference genome (or even to wild

stains), but it could also represent loss of elements in the reference genome or in wild strains.

>> We agree that, even with well established phylogeny and root, the two hypotheses are theoretically possible, and we discuss this point in the discussion L555-563, L641-643 .

- L285-288: Again, if the three cheese clades are ancestral (previous to domestication), there is no reason to suggest the observed structural variants are a respond of domestication.

>> Indeed, we may have missed some wild populations. We discuss this point in the discussion L555-563, L641-643.

- L503-507: "The genetic relationships between *G. candidum* populations and their contrasting levels of diversity suggest that domestication occurred in several steps, with an ancient domestication event separating the mixed-origin and the wild clades, then the cheese and the mixed-origin clades, and yet more recently the three cheese clusters.". It is not clear how much of the divergence between clades is due to domestication or already existing ancestral variation. For instance, the existence of the three-cheese clade within the cheese strains before domestication.

>> We have added a sentence in the discussion L641-643: "the divergence of the different cheese populations may have predated domestication; they would then correspond to unsampled wild populations, from which strains would have been independently isolated for cheese making". Nevertheless, as the three cheese populations displayed typical phenotypes of cheese fungi, it is likely that the three cheese populations are the result of domestication.

- L304: I was not convinced by the way authors identify genomic footprints of adaptation. They looked for regions of high Dxy between wild and domesticated populations and low diversity within populations as indication of regions under selection. But regions are selected from the distribution of variation along the genome. For instance, the 5% lowest divergent windows within the distribution. I do not think this is correct. Even under complete neutrality, variation in Dxy and diversity is expected, leaving to a distribution from which a 5% can be extracted.

>> We agree with the reviewer, we have now considered regions that pop up as extremes in both  $D_{xy}$  and pi scans as candidates for being involved in adaptation, and we ran SweeD tests.

- L331: No clear to me why it is used the mixed origin population. In the comparison of Dxy and Pi, the wild population is used. Here for the McDonald and Kreitman (MK) tests, the mixed origin is used instead.

>> In McDonald and Kreitman tests, we compare a focal group to the closest outgroup, here the mixed-origin clade, to infer non-synonymous

and synonymous changes taking into account mutation rates. For  $P_i$  and  $D_{xy}$ , using the closest outgroup is less critical, and we thought it would be better to compare to a wild population rather than a population in which some genes could already have been under selection for cheese making. Nevertheless, we also performed MK tests by using the wild clade as an outgroup to compare with the cheese and mixed-origin clades. We thereby detected 23 genes as evolving under positive selection, including a spermidine resistance protein, likely playing a role in yeast-hyphal transition. We have added this analysis in the paper L380-385.

- L28-29. L132-133. L512: "The domestication of *G. candidum* did not involve strong bottlenecks that occurred in other domesticated cheese fungi". I'm not sure there is a clear demographic analysis to make this claim. Diversity within populations is as low as in other domesticated fungi. An alternative hypothesis is that the bottleneck occurred in a admixed population with three main ancestral groups.

>> We have softened our claim about the absence of strong bottlenecks in cheese fungi. However, compared to other cheese fungi, the variability is higher, and most importantly, no *G. candidum* cluster is a single clonal lineage as found in *P. camemberti* and *P. roqueforti*.

- L534 or L33-34. "one of the cheese populations displayed footprints of a more advanced state of domestication". I found difficult to follow the discussion of a population being in a "more advance" or "least advance" state of domestication. From this paragraph I understand the connection is done based on difference in phenotypes observed between cheese associated and wild strains, as well as lowed genetic diversity in the former. However, these factors do not necessarily need to be due to domestication, thus I could not follow what is the reasoning to considering a population in a later or earlier stage of domestication.

>> We have tried to clarify the text and be less affirmative. Nevertheless, we find very interesting that a cheese clade has much lower diversity, phenotypes much closer to other cheese fungi, and much more differentiated phenotypically from wild strains. We agree that we do not know the ancestral state for sure, but these findings altogether do suggest that the Cheese\_2 population has a stronger domestication syndrome, which would be a pity not to discuss in our opinion.

#### **Minor comments:**

- L71-73. There are thousands of papers in fungi. Even in domesticated species. I argue this is not true.

>> A quick and rough search in Web of knowledge returns the following numbers of papers:

- "domestication and plant\*": 8,496

- "domestication and animal\*": 3,191

- "domestication and (fung\* or yeast\*)\*": 745 (i.e. 8% that of plants)

We have clarified our sentence, changing “little studied” into “less studied than plants and animals” (the sentence is about domestication), which we think is correct given the numbers above.

- For all phenotypic analyses. It was not clear to me if multiple test correction was applied. It was also not clear to me the number of replicates performed per strain. In Supplementary figures 8-10 shows N as the number of strains used, but it is not clear the number of replicates per strain. Is technical/methodological variation within strains included in the statistical model?

>> The Bonferroni correction was applied for the multiple test comparisons. We did not have replicates for our growth experiments as we aimed at testing differences between populations and not between strains: with a limited number of Petri dishes doable per experiment, the power is much stronger with more strains than more replicate per strains; the replicates only serve to have a finer estimate for the analysed growth parameters per strain as we do not test for a strain effect, only a population effect. We did perform some triplicates for lipolysis, proteolysis and volatile analyses, and used the mean for each strain to test for population differences to avoid pseudoreplication.

- There are a few instances in which the use of "recombination" is not used correctly, and it is used instead to refer to admixture event (recombination between divergent haplotypes). This needs to be reviewed along the manuscript.

>> We have replaced “recombination” by admixture when appropriate in the manuscript.

- L656: nbPROJECT need to be changed to the accession ID.

>> We have added the accession ID in the text L810.