Bennetot *et al*. studied patters of genome-wise 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 no 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 clams. I was also confused with some of the discussion points, which I believe should be clarified. Please see below my comments.

- The authors performed an admixture analyses 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 is 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.

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

- 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?

- 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.

- 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.

- 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.

- 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.

- 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.

- 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.

- 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 clam. 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.

- 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.

Minor comments:

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

- 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?

- 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.

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