



Towards a more accurate metabarcoding approach for studying fungal communities of fermented foods

Caroline Strub based on peer reviews by **Johannes Schweichhart** and 2 anonymous reviewers

Olivier Rué, Monika Coton, Eric Dugat-Bony, Kate Howell, Françoise Irlinger, Jean-Luc Legras, Valentin Loux, Elisa Michel, Jérôme Mounier, Cécile Neuvéglise, Delphine Sicard (2023) Comparison of metabarcoding taxonomic markers to describe fungal communities in fermented foods. bioRxiv, ver. 3, peer-reviewed and recommended by Peer Community in Microbiology. <https://doi.org/10.1101/2023.01.13.523754>

Submitted: 20 January 2023, Recommended: 29 August 2023

Cite this recommendation as:

Strub, C. (2023) Towards a more accurate metabarcoding approach for studying fungal communities of fermented foods. *Peer Community in Microbiology*, 100007. [10.24072/pci.microbiol.100007](https://doi.org/10.24072/pci.microbiol.100007)

Published: 29 August 2023

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Improved characterization of food microbial ecosystems, especially those fermented is key to the development of food sustainability. Short-read metabarcoding is one of the most popular ways to study microbial communities. However, this approach remains complex because of the locks and biases it may entail particularly when applied to fungal communities.

Building and using four mock communities from fermented food (bread, wine, cheese, fermented meat), Rué et al., 2023 demonstrate that combined DADA2 denoising algorithm followed to the FROGS tools gives a more accurate description of fungal communities compared to several commonly used bioinformatic workflows, dealing with all amplicon lengths. Moreover, Rué et al., 2023 provide guidance on which barcode to use (ITS1, ITS2, D1/D2 and RPB2), depending on the fermented food studied.

Practices in metabarcoding of fungi have been recently reviewed by Tedersoo et al., 2022 and their synthesis comes to the same conclusion as Rué et al., 2023. As the reference databases are far from being complete notably for food ecosystems, the development of specific sequences public databases will enable the scientific community to lift the veil on this whole area of microbial ecology.

The study conducted by Rué et al. (2023) provides a particularly detailed approach from a technical point of view, which contributes to improving the general practices in the metabarcoding of fungi. The design and the use of mock communities to compare the performances of the different pipelines is a strong point of this

study. Another key element is the creation and use of an in-house database of fungal barcode sequences which improved the species-level affiliations

However, the study of fungal communities by metabarcoding is still a promising avenue of research in agri-food sciences. Thus, short-read sequencing, combined with suitable pipelines and databases, should remain of interest to the microbial ecology community (Pauvert et al., 2019; Furneaux et al., 2021).

References:

Furneaux, B., Bahram, M., Rosling, A., Yorou, N. S., & Ryberg, M. (2021). Long-and short-read metabarcoding technologies reveal similar spatiotemporal structures in fungal communities. *Molecular Ecology Resources*, 21(6), 1833-1849. <https://doi.org/10.1111/1755-0998.13387>

Pauvert, C., Buée, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., ... & Vacher, C. (2019). Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecology*, 41, 23-33. <https://doi.org/10.1016/j.funeco.2019.03.005>

Rué, O., Coton, M., Dugat-Bony, E., Howell, K., Irlinger, F., Legras, J. L., ... & Sicard, D. (2023). Comparison of metabarcoding taxonomic markers to describe fungal communities in fermented foods. *BioRxiv*, 2023-0113.523754, ver.3 peer-reviewed and recommended by Peer Community in Microbiology. <https://doi.org/10.1101/2023.01.13.523754>

Tedersoo, L., Bahram, M., Zinger, L., Nilsson, R. H., Kennedy, P. G., Yang, T., ... & Mikryukov, V. (2022). Best practices in metabarcoding of fungi: From experimental design to results. *Molecular ecology*, 31(10), 2769-2795. <https://doi.org/10.1111/mec.16460>

Reviews

Evaluation round #2

DOI or URL of the preprint: <https://doi.org/10.1101/2023.01.13.523754>

Version of the preprint: 2

Authors' reply, 08 August 2023

Dear recommender,

We have answered all reviewer comments and hope this version will meet PCI requirements.

My best regards

Delphine Sicard

[Download author's reply](#)

[Download tracked changes file](#)

Decision by [Caroline Strub](#), posted 02 August 2023, validated 02 August 2023

One last adjustment

Dear authors,

Could you consider the minor comments, in particular the one about methods (line 330 : Is it an OTU represented

by a centroid, a Swarm seed or a denoised sequence variant ?) and the manuscript will be ready to be recommended.

Sincerely,
Caroline Strub

Reviewed by **Johannes Schweichhart**, 01 August 2023

I only have a few minor comments and one issue which has been addressed before and is not resolved yet. Otherwise I would recommend this preprint for publication.

Ad Introduction:

Line 81: This sentence does not really reflect the findings of the Ihrmarks paper and contradicts the findings in the preprint which shows a high divergence rate for all pipelines.

L114: I guess "downside" not "downfall" is meant here.

L124: Building correct biological sequences is beside the point of traditional de novo clustering.

Ad Methods:

L330: I thank the authors for their answer but apparently no changes were made in this respect in the preprints methods. To be more explicit: A difference that can make a lot of difference, especially when talking about perfect matches to reference sequences, is what is compared with that reference sequence - is it an OTU represented by a centroid, a Swarm seed or a denoised sequence variant? This is not implicit for every pipeline and "following authors guidelines" is too unspecific for USEARCH and Qiime. I guess for USEARCH the authors refer to "recommended procedures" at <https://drive5.com/usearch/manual/>. There both, OTU clustering and denoising, are given which makes this reference ambiguous on how things have been done in the preprint. Similar is true for Qiime. It should not be necessary for the reader to screen the code in the supplementary just to get the information if the respective pipeline was using ZOTUs, ASVs or OTUs.

Ad Discussion:

L680: It is rather likely that all primers have mismatches with certain groups of fungi.

Reviewed by anonymous reviewer 2, 02 August 2023

The amended version and the author's responses are satisfactory.

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.1101/2023.01.13.523754>

Version of the preprint: 1

Authors' reply, 05 July 2023

[Download author's reply](#)

[Download tracked changes file](#)

Decision by **Caroline Strub**, posted 19 June 2023, validated 19 June 2023

Moderate revision

Dear Authors,

I would like to apologize for the delay in the handling process of your preprint.

You present a novel approach to solve the issue concerning the length polymorphism of ITS1 and ITS2 sequences in metabarcoding of fungi.

Both reviewers and I agree this is a relevant study which requires moderate revision, following comments by the reviewers.

Sincerely,
Caroline Strub

Reviewed by anonymous reviewer 1, 04 April 2023

This study entitled "Comparison of metabarcoding taxonomic markets to describe fungal communities in fermented food" can be divided into two sub-sections: First a comparison of mock communities of four common fungal identification markers (ITS1, ITS2, D1/D2, and RPB2) and seven bioinformatics workflows (using four different bioinformatics tools) including the most common approaches (OUT, ASV, ZOTUs) with a focus on fermented foods using four fermented food models (bread, wine, cheese, fermented meat).

The title reflects the content of the paper and the main results of the study are summarised in the abstract. The research question is very relevant to the field of food microbiology and microbial ecology and is well addressed, using relevant approaches and tools.

This paper provides an excellent contribution to the field of food microbiology. The authors demonstrate a thorough understanding of the subject matter and present a well-designed study that compares different metabarcoding pipelines and markers to determine fungal diversity in food ecosystems, with a focus on fermented foods.

The use of mock communities to validate the bioinformatics tools is a particularly strong aspect of this study, as it allows for rigorous testing of the pipelines in a controlled setting.

The comparison of four bioinformatics pipelines, including DADA2, QIIME, FROGS, and a combination of DADA2 and FROGS, is also noteworthy. The authors' demonstration of the superiority of the combined DADA2 and FROGS tools will be of interest to researchers in this field.

The paper highlights the importance of selecting appropriate markers, with the authors finding that ITS markers performed better than D1D2. The study provides guidance on the best markers for different food ecosystems, with ITS2 being best suited to characterize cheese, wine, and fermented meat communities, while ITS1 performs better for sourdough bread communities.

Overall, this scientific paper presents a thorough and well-executed study that makes a valuable contribution to the field of bioinformatics. The questions addressed are relevant to the field and the results will be of interest to researchers in agri-food sciences and microbial ecology, and the paper provides a framework for future research in this area.

I therefore recommend this paper for submission.

I also have some specific comments for the author to address:

- The relevance of figure 1 is questionable. It does not bring a substantial amount of information. Moreover, data in the text below figure 1 do not seem to confirm the data represented in figure 1: for example for meat, the text states that 4 species per genus were found for *Yarrowia* and *Cladosporium* and 2 species for *Candida* only. However, in figure 1, only a single dot can be found at 4 species/genus and 3 dots at 2 species/genera. Either there's a consistency problem between the figure and the text (especially for meat) or the data are confusingly expressed.

- Species names should be italicized (lines 249, 250, 455, 513)
- The ITS region being subjected to significant size polymorphisms (insertion/deletions) as shown in fig2 and fig 6, It is often difficult to interpret/make sense of phylogenetic trees built on sequence alignments. It might be interesting for the authors to elaborate and discuss on the relevance of the trees obtained and shown in fig3
- Figure 5:

- Panels should be numbered or labeled (A, B, C, D)
- Precision should be provided in the legend to clarify the difference between the small dots and the bigger dots
- Figure 6:
 - Figure legend stats "ITS1 and ITS2 amplicon size...". However, it seems that only ITS1 data are presented.
- Figure 8:
 - chosen colors make it difficult to distinguish partially reconstructed and perfectly reconstructed sequences. A better choice of colors would greatly benefit the readability of the figure

Reviewed by [Johannes Schweichhart](#), 18 June 2023

Provide a detailed, objective report on the merits of the preprint.

- Rue et al. benchmark combinations of four barcoding marker regions (ITS1, ITS2, RBP2 and D1/D2 LSU) and bioinformatic pipelines (USEARCH, QIIME2, DADA2 and FROGS) in terms of their capability to recover fungal species and type sequences from four mock communities for fermented meat, wine, cheese and sourdough (118 fungal species in total). Based on merged data from all four marker regions the authors conclude that a combination of FROGS procedures and the DADA2 denoising algorithm shows better performance than the other pipelines tested. They then apply this procedure to fermented meat, wine, cheese and sourdough samples (n=24) and provide a description of their findings.

Identify flaws (if any) in the design of the research, and in the analysis and interpretation of results.

- Ad Fig. 3: To my knowledge, phylogenetic relationships derived from ITS1 and ITS2 sequences become unreliable above class level. If the authors want to include this figure, I would suggest to use LSU sequences instead.
- Ad "Analysis of real samples" (lines 449 ff.): Results are only discussed in the context of marker choice. This might be misleading. In the case of missing *Yarrowia* and *Candida* species when using the ITS1 marker, this can more likely be attributed to the primers used (ITS1F and ITS2) which have known mismatches to those taxa (see [Tedersoo and Lindahl, 2016](#)).

Expose your concerns (if any) about ethics or scientific misconduct.

- No concerns.

State the preprint's strengths as well as its weaknesses. Try to consider both the technical merit and the scientific significance.

- An unsolved issue in short-read metabarcoding of fungi is the length polymorphism of ITS1 and ITS2 sequences which commonly leads to the exclusion of taxa with longer variants of those markers. The authors of FROGS introduce a novel approach to solve this issue and, based on simulated data, show that this leads to higher recovery rates of fungal taxa without compromising on precision in the taxonomic classification of the fungal community. In this preprint, this approach is extended to four barcoding marker sites (of which one is out of range for the short-read sequencing platforms) and four fungal mock communities of fermented foods. While not completely novel to this preprint it is a promising methodological approach which has potential to improve the general practice in the metabarcoding of fungi and should be promoted.

- The preprint is very detailed on the technical side of things but rather brief concerning the biological backgrounds and the presentation of the results for real food samples. Thus, the motivation for this study is not very clear. The comparison of the results of fermented food samples is purely descriptive and limited to a few selected findings which are only attributed to targeted marker sites. As mentioned briefly by the authors, primer bias is another important factor and should be included. A graphical representation of these results is missing.

If there is something critically missing, report it.

- Lines 125 ff.: A key component of this study are fungal mock communities representative for fermented meat, wine, cheese and sourdough. The authors write that the selection of those species was "based on an inventory of the most frequently described species in the literature". Yet, neither references for the literature sources used for that purpose nor further methodological insight how these species were selected are given.
- Lines 294 ff.: Which algorithms were used here? UPARSE or UNOISE (i.e. OTU- or ZOTU delineation)? If UPARSE the comparison of perfectly reconstructed sequences (Fig. 5) would not be sound since OTUs would be compared to ASVs. Similarly, was QIIME2 used with the DADA2 plugin or using a OTU clustering algorithm?

Provide specific suggestions for improvements.

- Ad Fig. 5: In my opinion, the outcomes of the benchmark of the pipelines used would be more clear and transparent if the results from all four marker sites would be presented separately and not mixed together.
- Consider re-running some of the analyses using the current version of UNITE (v9.0). The version used here is >2 years old and since then ~9x more fungal sequences and >25% more fungal reference sequences have been added to the database.
- Colours with better contrast could be picked in the Fig. 8 to discriminate partially and perfectly reconstructed sequences.