- **Title and abstract**
    - Does the title clearly reflect the content of the article? [ ] Yes, [ ] No (please explain), [ ] I don't know
    - Does the abstract present the main findings of the study? [ ] Yes, [ ] No (please explain), [ ] I don't know
- **Introduction**
    - Are the research questions/hypotheses/predictions clearly presented? **[ ] Yes,** [ ] No (please explain), [ ] I don't know
    - Does the introduction build on relevant research in the field? [**] Yes,** [ ] No (please explain), [ ] I don't know
- **Materials and methods**
    - Are the methods and analyses sufficiently detailed to allow replication by other researchers? [ ] **Yes,** [ ] No (please explain), [ ] I don't know
    - Are the methods and statistical analyses appropriate and well described? [**] Yes,** [ ] No (please explain), [ ] I don't know
- **Results**
    - In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? [ ] Yes, [ ] No (please explain), [ ] I don't know
    - Are the results described and interpreted correctly? **[ ] Yes,** [ ] No (please explain), [ ] I don't know
- **Discussion**
    - Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? **[ ] Yes,** [ ] No (please explain), [ ] I don't know
    - Are the conclusions adequately supported by the results (without overstating the implications of the findings)? [ ] Yes, [ ] **No (please explain),** [ ] I don't know
        - I think there should be a bit more of this, as mentioned in my comments. Especially that there is little to no discussion of MQ and Mascot, the most frequently used programs in ancient proteomics. Also, I would add more detail and a figure to show what an ideal data analysis pipeline would be

Line 223 - Dairy database

Bleasdale 2021 and Wilkin 2021 did not look for only dairy sequences, the goal was to identify any dietary proteins from consumption. The samples were searched against Swissprot entirely and also included a custom dairy database as well.

Line 341 “less accepted PSMs” should be “fewer accepted PSMs”

Lines 318-321
Can you explain what the q_value is telling us? Is this like the expect value of the peptide ID? Or is this something related to FDR?

Lines 334-336. If there is no examination of the quality of the IDs then is there a point of comparing them? Maybe I’m missing something, but if FragPipe is making a ton of IDs but they are unreliable or not found through other searches, do they really count? Are the IDs too good to be true? I’m honestly asking, as this has been a large concern of mine when deciding which program to use.

LIine 380 - Does “Figure 2” mean Figure 3? Also, I would reiterate in the legend to Figure 3 which are narrow and which are open again. Otherwise readers may have to flip around the paper to remind themselves which is which.

Line 397 - Referring to “Figure 3” is actually Figure 4 (with Venn diagrams). Check all figure mentions throughout, as many are incorrect.

Figure 4 - can you add a total of all PSMs recovered from each entire Venn diagram? This would help rather than readers adding them up.

Figure 5. What are the shaded regions underneath on the animo acid counts? This is mentioned in the main text at different points, but should also be referred to in the figure legend.

Overall-
I am missing any discussion of the narrow searches. It would be great to have some assessment of how MQ and Mascot perform as these are the most commonly used programs in ancient proteomics. Are these less reliable than the open searches? They do seem to ID fewer peptides, but are the open search IDs of good quality? If not within the scope of this paper, this would be a great next step.

In conclusion, I would make a stronger stance on how to best go about searches. Maybe some sort of table or figure that outlines the suggestions of using an open search first and then narrowing the space.