



**MSc Bioinformatics
Dissertation Report
Marking Feedback**

Student Number:	B269797
Project Title: (as per original database entry)	Metagenomic insights into sourdough-based Quas fermentation and dietary fibre addition

Marking Criteria	Marker 1 Grade	Marker 2 Grade
Acknowledgements (optional)	B	A3
Abstract/summary	A3/B	A3/B
Introduction	B	A2
Methodology	A3/B	A3/B
Results	B/C	A3/B
Discussion	B/C	B
Conclusion	A3	A3
Bibliography and references	A3	A3
Supplementary data/information/files	B	A3

1st Marker Feedback

Abstract/summary: Very high quality, providing a succinct overview of the study and the main findings, but it is quite technical in places.

Introduction: Very good on the whole, providing an overview of Kvass and biotic communities, and how such inocula are investigated using sequencing methods. I particularly liked the end of the section, which outlined explicit hypotheses tests and overall aims of the project. This section could have done with some editing, as there were several sections that were less relevant (e.g., on the proposed health benefits of beverage microbial communities), and some technical terms could have been defined (e.g., 'alpha' diversity, 'V3–V4' regions).

Methodology: Excellent overall, providing the right level of detail when describing the DNA extraction methods, quality control, and subsequent analysis. I particularly liked the use of Figure 1 to describe the workflow. Towards the end there was an introduction of several computer methods (e.g., Kaiju, DRAM) but it was not clearly explained what each of these programs do.

Results, Discussion: This section is generally competent, describing the composition levels of each test; the diversity present in different cultures; and gene ontology analysis for each community composition. The hypotheses tests are left for the discussion, but I found that using this strict separation made it harder to fully interpret the results. I had to flick back-and-forth to see which results matched up with each hypothesis test, limiting my ability to fully understand the scientific implications of the work. I also found the heatmap plots (and the resulting discussion) very abstract and hard to interpret (for example, it was not made clear what a 'CAZy' family is, and some of the statistics used to measure alpha and beta diversity were not fully explained). Good ideas for future research are proposed.

Conclusion: Excellent summary of the project.



Bibliography and references: Well used throughout.

2nd Marker Feedback

Abstract/summary: generally good, but the reader is left without really being told which, if any, formulations would be better - only that they are different.

Introduction: The introduction is excellent - nicely written and structured - and it gives a nice overview of what is known. Some bits, however, come across as a bit jargony and AI-like (for example, the significance section at the end of the introduction).

Methodology: The methodology is comprehensive, and a reader who wishes to replicate the study could do so, I believe, from what is written. While I realise this is standard in bioinformatics, I always find it frustrating when the methods contain a list of software without any real indication of how the method works. For example, it is stated that DADA2 denoising was performed - but I have no idea what this means. The section on diversity was unclear - I was expecting a single value for each formulation, but then the results contain boxplots and PCA scores. Presumably these metrics are calculated for some taxonomic grouping? In addition, the motivation for using the different metrics and their interpretation were not clear.

Results: The results are generally well presented, although I think they could be improved with a little interpretation (see above regarding the diversity indices).

Discussion/Conclusion: This was probably the weakest part of the thesis, as the discussion often felt like a rerun of the results. It would have been nicer to add more interpretation and be clearer about which, if any, of the formulations have advantageous consumer/health properties. The single-sample setup is essentially fatal—without any information about between-batch variability, there is no way to know if the patterns described have occurred because of the different formulations or by chance. I have not taken this into account in the mark—this is a failure in study design, not the student.

Bibliography and references: well referenced throughout.

Supervisor Feedback

XX demonstrated excellent performance throughout this project. She is very self-directed, technically excellent, and capable of independent troubleshooting. She showed a great ability to learn quickly and effectively incorporate feedback into her work, as evidenced by the substantial improvements made to her final dissertation. While there was an initial over-reliance on AI tools for writing, she took this feedback on board and successfully revised her work to reflect her own strong analytical abilities. Her dedication and competence have resulted in a high-quality piece of research. Well done.