

In this assessment, you will be provided with paired-end short read sequence data (coursework2023\_R1.fastq and coursework2023\_R2.fastq). The assessment requires you to analyse these data and answer a series of questions. In your answers, make sure to outline the approach taken and, where appropriate, the evidence for your conclusions. The suggested word limits are indicative, and you can write less than this if you provide a full answer to the question. Screenshots or plots can be included to illustrate the results of your analyses and support your arguments.

1. (a) Evaluate the quality of the raw sequencing data using FastQC (10%).  
  
(b) Trim the reads to remove adapter sequences and low-quality bases. Assemble the reads and indicate the total length of the assembled contigs, the N50 and the mean coverage (10%).  
  
(c) Write a Methods section for the protocol used in the style of the journal *Microbial Genomics* (Up to 300 words, 20%).
2. How many organisms were sequenced in this sample? State the name of each species identified. Summarise the evidence for your answer (Up to 200 words, 10%).
3. Propose and justify, with evidence from your analyses, a hypothesis for the core energy metabolism of each of the predominant community members (Up to 300 words, 30%).
4. The species in this sample have proven to be very challenging to grow in isolation (that is, as a pure culture). Propose, providing evidence, one reason why this might be the case. Your evidence might come from the metabolic inferences in (3), from your reading of the literature, or from another source (Up to 300 words, 20%).