**Name**:

1. The **FastQC** step
   1. Paste the screenshot of your complete MultiQC \*.out file (alternatively, the result of listing your fastQC directory using *ls -lh*). Make sure the text in the screenshot is legible when enlarged (the same below for all screenshots). (Note: step 1 of your HW2 will not be graded if this screenshot is missing or illegible)
   2. The smallest and largest library sizes (Millions of fragments), respectively, are:
   3. Paste the screenshot of the Per Sequence Quality Score graph generated by MultiQC in multiqc\_report.html (alternatively, the individual screenshots of the corresponding 30 graphs generated by fastQC)

2. The **fastp** step

1. Paste the screenshot of the bash command part only of your fastp script, including the bash command sequence and the for loop
2. Report the fastp processing statistics of SRR8353425 (should be at the very bottom of your \*.out file)

|  |  |
| --- | --- |
| Items | Numbers (3 significant figures) |
| reads passed/after filter (millions) |  |
| reads failed due to too short (millions) |  |
| duplication rate (%) |  |

3. The **STAR** step

1. Paste the screenshot of your complete MultiQC \*.out file (alternatively, the result of listing your STAR directory using *ls -lh*). (Note: step 3 of your HW2 will not be graded if this screenshot is missing or illegible)
2. Paste the screenshot of the “General Statistics” table generated by MultiQC in multiqc\_report.html (alternatively, manually create a table showing the Uniquely mapped reads (% and number) for each SRR from its \*.Log.final.out file). Note: Sort the % Aligned column from the highest to the lowest.

4. The **featureCounts** step

1. Paste the screenshot of your complete MultiQC \*.out file (alternatively, the result of listing your featureCounts directory using *ls -lh*). (Note: step 4 of your HW2 will not be graded if this screenshot is missing or illegible)
2. Paste the screenshot of the “General Statistics” table generated by MultiQC in multiqc\_report.html (alternatively, manually create a table showing the Assigned reads (% and number) as you can get from the featureCounts\_PE.txt.summary). Note: Sort the % Assigned column from the highest to the lowest
3. In the generated count table, the counts of gene ENSRNOG00000016227 in SRR8353406 =

5. **Bonus** questions ( 3 pts)

a. Why the duplication rate in SRR8353425 appears much higher in the FastQC report than in the fastp report? Answer briefly using a maximum of 2-3 sentences (hint: do an online search). ( 1 pt)

b. For reads that are each mapped to multiple (*n*) genome loci and thus cannot be uniquely assigned, what is the average value of *n* in SRR8353396? (hint: consider and compare the multiple loci mapping values in STAR and featureCounts reports) (2 pts)