NGS1 Worksheet
Name:

Instructions: Instructor will randomly select one copy of the worksheet from your group at the end of the lab. Make sure to answer questions completely and clearly, fully indicating your reasoning/thoughts (often writing in complete sentences is helpful). I expect the responses to **have no typos and minimal grammatical error**. You are allowed and encouraged to discuss questions with your labmates.

A1. For the question: "Are microbial communities found on an individual organism consistent across body parts?", form a hypothesis.

B1. What date and time is associated with the 'data' directory that Dr. Yang made?

B2. Fill in the following table.
Table 1. Metrics for the two sequencing read datasets available for female garter snakes for mouth glands (oral2) and musk glands (musk2)

Filename	female_oral2	female_musk2
a. Size of Zipped File (Kilobytes)		
b. Size of Unzipped File (Kilobytes)		
c. # of lines in file		
d. # of reads per file		

B3. Explain what each line of the first four lines show in a FASTQ file.

C1. In the female_oral2.fastq file, the first read's quality score for the first base is 'G'. Explain what this quality score means (numerically and conceptually).

C2. In *fastqe*, what options allow you to examine the maximum and minimum read quality scores for each position?

C3. Fill in the following table.

Table 2. Metrics for the two sequencing read datasets available for female garter snakes for mouth glands (oral2) and musk glands (musk2) after filtering out low quality reads

Filename	female_oral2.q15	female_musk2.q15
a. # of lines in file		
b. # of reads per file		
c. # of reads filtered out		
d. Examine the table and two graphs specified in step C10 and describe the differences in read quality before and after filtering out low quality reads using fastp (e.g. % reads passing filters, Q20/Q30 changes, describing change in graphs from before to after filtering). Try to be specific - use numbers to help convey differences you notice.		

C4. We used the default filters (see step C6). What do you think the advantages are for using a more stringent filter? What about disadvantages? You do not need to run anything, but if you'd like to try it out to check your reasoning, use the option -q # to adjust the filter. For instance, a more stringent filter you could try is fastp -i female_oral2.fastq -o female_oral2.q30.fastq -q 30

C5. Which tool, *fastqe* or *fastp*, did you find easier to use? Explain your reasoning.

C6. Which tool, *fastqe* or *fastp*, do you think is a more reliable research tool? Explain your reasoning.