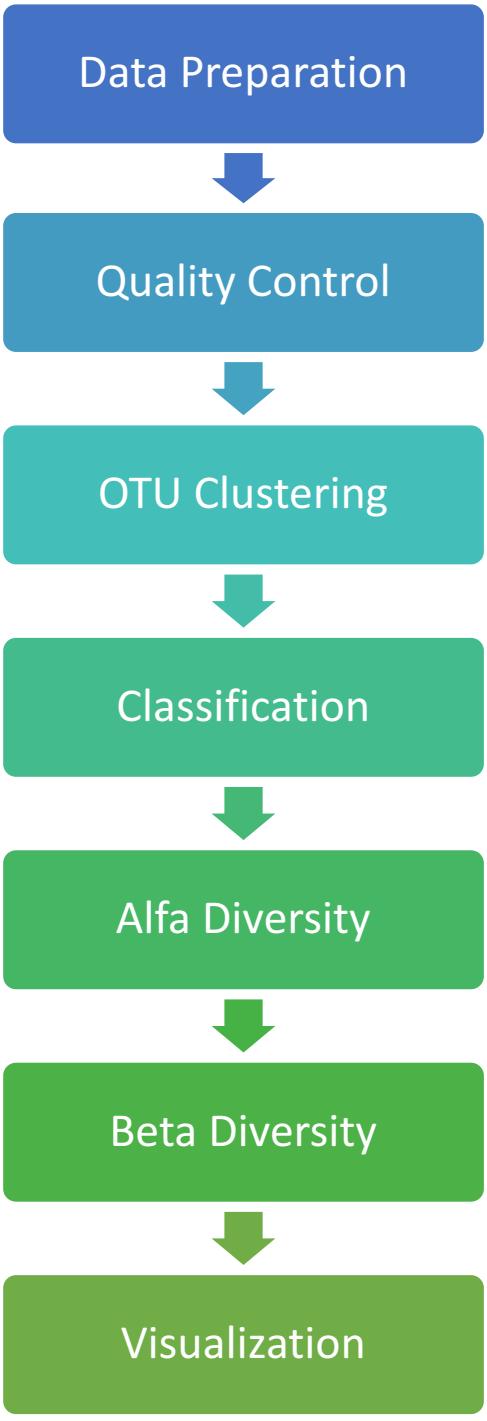
Student: _____

PRACTICAL – EXIT TICKET

We applied this workflow in the practical. Fill it with the format file and software used in each step:



GALAXY TOOLS: List of Dataset Pairs

GALAXY TOOLS: Screen.seqs

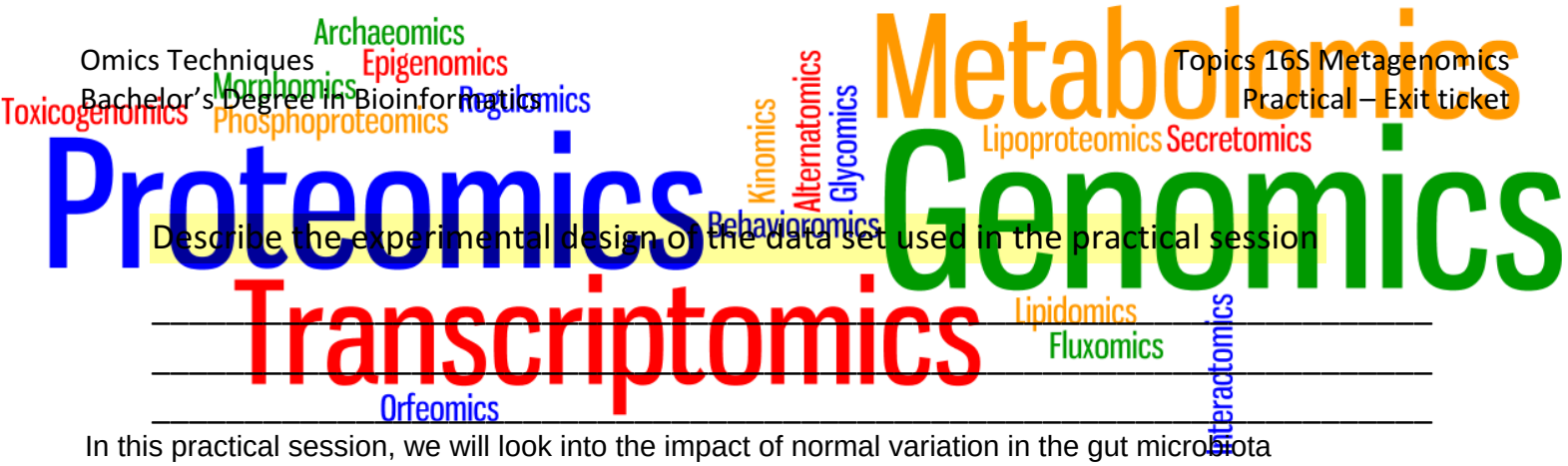
GALAXY TOOLS: Cluster.split

GALAXY TOOLS: Classify.otu

GALAXY TOOLS: Rarefaction.single

GALAXY TOOLS: Dist.shared

GALAXY TOOLS: Make.biom



Describe the experimental design of the data set used in the practical session

In this practical session, we will look into the impact of normal variation in the gut microbiota on host health. By testing mice's feces on a regular basis and determining if the quick change in weight altered the stable microbiota compared to the microbiome observed in the previous days.

How does the duplicate sequences removal works?

Microbiome samples often include huge numbers of the same organism, thus we anticipate to discover many similar sequences in our data. In this situation, we will use the Unique.seqs tool to delete the identical sequences by first determining the unique reads and then recording how many times they exist.

Why is the OTU clustering one of the main steps on the analysis? What would you expect (in terms of diversity) if we apply an identity threshold < 97%?

It is an important step because OTUs are clusters of similar sequence variations of the 16s rDNA marker gene sequence, each of which is intended to represent a taxonomic unit of bacterium sequences or genus (depending on the sequence similarity threshold). We use the 97% because we wish to discriminate at the species level.

Explain the importance of rarefaction curves. If you have a sample which show a rarefaction curve which has not started to level off, how would you solve this problem? Give at least two possible solutions

Rarefaction curves show the number of species as a function of the number of individuals sampled. If we have additional funds to sequence, we can add more duplicates to a sample that has not yet plateaued. If not, we may search for additional studies that have sampled from this setting to supplement our findings. Still, this is not an incorrect finding; however, we must note that our results show a partial depiction of the curve.
We may use statistical distributions to analyze it. We might use either the logarithmic distribution or the broken stick distribution.