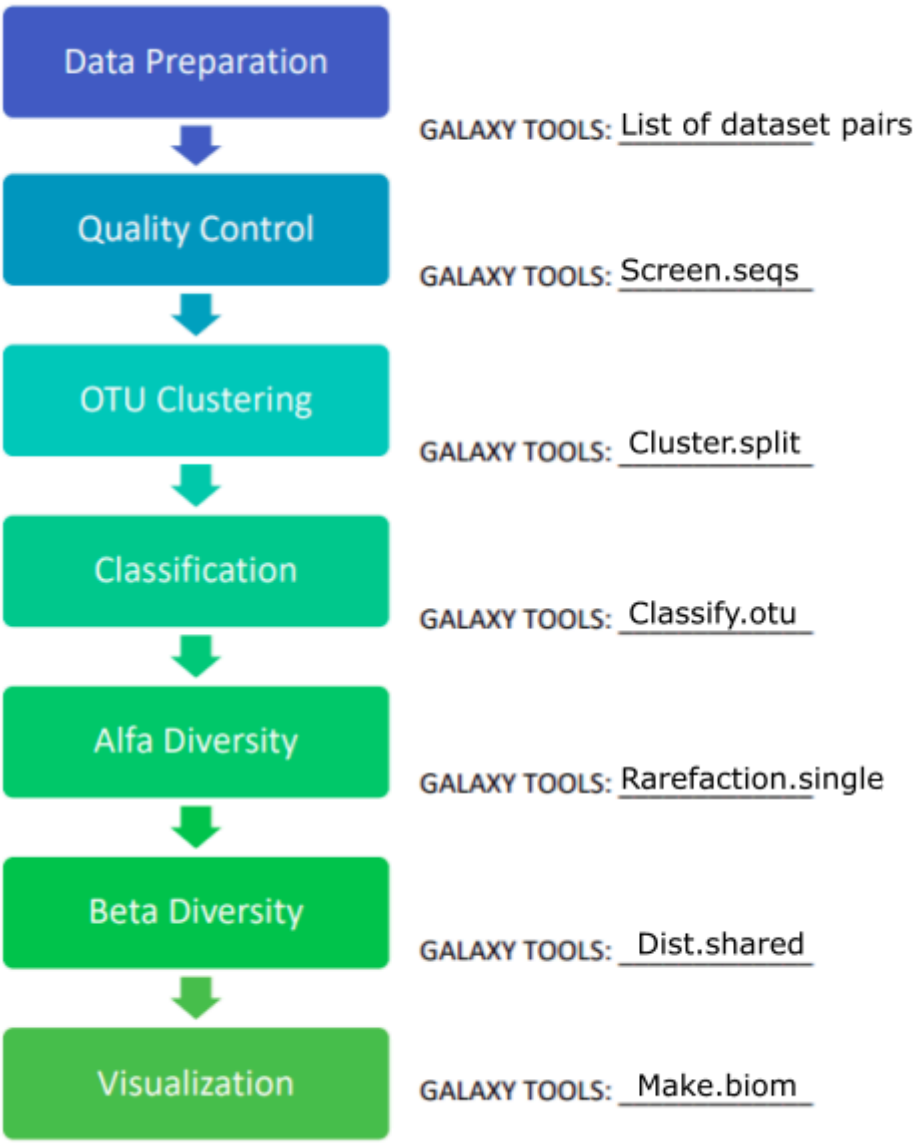
PRACTICAL EXIT TICKET

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



Jan Carreras Boada 107765

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

Over a 365-day period, mice's feces were collected to study changes in the gut microbiome. To ensure accuracy, we employed a mock community with a predetermined composition and analyzed data from a single mouse at different intervals.

Each sample is represented by a file for both its forward and reverse strands, for a total of 40 paired fastq files containing the collected data. We are going to use the Galaxy platform to process this data.

How does the duplicate sequences removal works?

We first identify the reads, count how many times they occur, and then optimize the files for computational efficiency, which improves processing speed, to remove duplicate sequences. We anticipate many identical sequences in microbiome samples, so it is imperative to eliminate duplicates.

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%?

OTU clustering is a critical method for taxonomic profiling because it facilitates the identification and classification of organisms. Additionally, OTU clustering reduces data volume by grouping 16S rRNA reads into clusters, making it easier to handle large datasets.

Setting an identity threshold below 97% would likely result in the observation of increased diversity. This occurs because a lower threshold results in the creation of more clusters, each potentially representing a distinct species.

Jan Carreras Boada 107765

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 are used to evaluate the adequacy of sampling efforts and to estimate species richness. They are essential for determining how comprehensive assessments of biodiversity are.

A rarefaction curve that does not exhibit a leveling off point may indicate that the biodiversity sample is not representative enough. Two possible methods to deal with this are:

1. Increasing the depth of sequencing or gathering more samples in order to obtain more thorough data.

2. Making use of subsampling techniques, which help to lessen the dominance of abundant species by randomly choosing subsets of individuals from samples to create multiple subsamples.

Reaching a plateau on a rarefaction curve signifies that the diversity of the majority of species has been sufficiently captured. This plateau indicates that a significant increase in species identification is unlikely to result from additional sampling.