## Task 2 : RNA Sequences :

The first step for this task is initialize all the constant we will use to realize this task. To do so, we redefine the data folder we are going to use, define the header for the columns from the experiment file. So now we have two columns header named ‘Name’ which represent the name of the bacteria and ‘Number’ which is the result of the experiment. Finally we initialize a dataFrame named ‘mb‘ which contains our final result.

In a second time, we load the metadata file to complete the header we initialize before. To perform this, we just add the first line of the metadata.xls file to the variable which contain the header I initialize before. Now we have the header of our result dataFrame.

Now we can begin to load the result of the experiments and store the in the same dataFrame. In order to do that, we assume that each experiment file are respect the following regular expression MID + [0-9]\* + .xls. This allow us to loop in the microbiome folder and load each file in the loop. Now we got the data from a single file loaded in a dataFrame, we set the column of this dataFrame to the ‘Name’ and ‘Number and then we add for each row the metadata related. Finally we add the data from the single file into the ‘mb’ dataFrame.

To complete the task we just set the index of ‘mb’ to ‘BARCODE’ so the data are indexed by experience number and we replace the NaN value by unknown. To visualize the result we save ‘mb’ into a .csv.