**Data files: UniProt2PlantReactome-pathways.txt and UniProt2PlantReactome-reaction.txt**  
**Students will merge data from these two files so that each UniProtID has the mapping to pathways and reactions.**

1) A simple join performed on the two datafiles based on the common column “UniProt.ID”. We got 32871 observations across 7 variables denoting pathways and reactions. The file is named as “UniProtID\_merged.rmd”.

2) However, this process, gives a lot of redundancies of rows, which we are focusing to reduce.

3) There are 251 labels for pathway names, many of which are reductant labels/sub labels (referring to gramene website), which we have to remove.

3) We only plan to retain the “Metabolic Reactions”. To do this we perform the following:

1. Removed rows having Pathway names matching “Translation elongation”.
2. Removed rows having Pathway names matching “Translation termination”.
3. Removed rows having Pathway names matching “Ribosome”.
4. Removed rows having Pathway names matching “subunits”.

-----> Reduced to 18057 observations