

Exercise (Introduction to Linux)

For each of these questions, provide the commands you used: (prepare an R markdown document)

1. Create a project directory called Exercise. Which commands did you use?
2. In the directory, create all the sub-directories you would need for a Bioinformatics Project
3. With the provided dataset files, put them in the appropriate directories you created under your Bioinformatics Project
4. Extract the sequence headers and save into a file `sequence_names.txt` in the appropriate directory
5. Save the commands you used in question 4 in a script file `extract_seq.sh`
6. Count the number of mRNA.
 - What other sequences apart from mRNA?
 - For each category, how many are there?
 - Save the above output to files.
7. How many organisms (create a file with the organisms without duplicates)
8. How many are predicted?
9. How many nucleotides are in the file? How many of each of the bases are there?