1. **Project work/Project Scope (As included in the Coursework projects.pdf located in the 1. Requirements and useful Project Info folder)**

Context: you want to develop a DNA identification service

• Identify the closest sequence in the database to the provided sequence – file: mystery.fa

• Input: sequence database (adapted from GEO), test sequence – file: dog\_breeds.fa

• Output: the closest sequence, and the difference

• Stretch goal 1: Probabilities across database, p-value

• Stretch goal 2: reconstructed phylogeny

1. **Project guidance/structure (As included in the Coursework projects.pdf, located in the 1. Requirements and useful Project Info folder)**

• Write Python code that reads in some input, does some processing, then output some results.

• This should take the form of a complete project:

• separate data files

• An executable python script

• Generation of an output

• Commented and tested

• With version control used

1. **The project should contain: (As detailed in the coursework\_checkpoint.pdf file located in the 1. Requirements and useful Project Info folder)**

• A README file

* A project structure

• Folders containing data, code, results, etc…

• Files containing the code, the data, etc…

• Preferably, stored in subfolders

• Make a folder, and get started!

• Making a plan

* You start with data: matrices, sequences, …
* You need to provide an output: text, graphs, etc…
* Split the whole project into steps. What steps are needed to go from one to the other?