

Sequence Classification

The zip file that comes with this project has five files. Four of them, synd1.fasta-synd4.fasta contains examples of different types of DNA. The fourth **TestData.fasta** contains sequences from all four categories, but different ones from the example files. The project asks you to create code that, given the sequences in TestData.fasta, correctly identify which category they belong to.

Your project must be coded in Python or R, your choice, and provide as output the category numbers (1-4) of the sequences in the TestData.fasta file. You must explain what method you are using. Any available python or R package may be used to solve the problem, but please document what you do.

Be sure to validate the input of your sequences - these fasta files use more than 80 characters per line. The entire example sequence appears on one line.