

Python Exercises - DNA Sequence Manipulation

(modified from Mayo Röttger and Andrea Schrader)

Programs should be written in Python according to the respective exercise description. The program must be correct in terms of syntax and semantic. If there exists a minimal solution using only a single pre-defined function, this function is not allowed. The weekly exercise should be uploaded on ILIAS as a single Jupyter Notebook (.ipynb). The assignment will be provided after the lecture on Wednesday and must be uploaded by Tuesday 8:00. Make sure that answers to questions are contained within the Jupyter Notebook that is uploaded on ILIAS. Use commented code when possible

1. Wheat Crop Yield Ratios

Write a script that takes the yields of three different wheat crops and calculates the ratios between them. The result should be printed, displaying at most 2 decimal places.

Task:

1. Prompt the user to input the yields of three different wheat crops.
 2. Calculate the ratio between the yields.
 3. Use f-string formatting to print the results with two decimal places.
-

2. Finding Substrings in DNA

Create a function that takes a long DNA string (100 characters) and finds the first occurrence of a specified substring. Test the function with two cases: one where the substring is present (ATCG) and one where it is not (AAAAAA).

Task:

1. Write a function that finds the first occurrence of a substring in a DNA sequence.
2. Test the function using ATCG (which is present) and AAAAAA (which is absent).

Use this DNA string for testing:

```
GCTAGTCGATCGTAGCTAGCTACGATCGTAGCATCGATCGTTAGCTAGCGTACGATCGTGCTAGCGT
```

3. Reverse Complement of a DNA Sequence

Write a function that creates the reverse complement of a DNA sequence. The function should reverse the sequence first and then replace nucleotides as follows: A -> T, T -> A, C -> G, G -> C. Hint python can distinguish between uppercase and lowercase letters.

Task:

1. Reverse the DNA sequence.

2. Replace nucleotides with their complement.
-

4. Printing DNA Before and After a Substring

Create a function that prints the DNA sequence before and after the occurrence of a given substring. Use the `find_substring()` function from Exercise 2 to locate the substring.

Task:

1. Use the function from Exercise 2 to find the position of the substring.
 2. Print the DNA sequence before the substring and from the end of the substring to the end of the sequence.
-

5. Searching for a Substring in Both Forward and Reverse Strands

Create a script that reads a DNA string from the user, checks if a substring is in the forward strand, checks if it's in the reverse strand, and prints the nucleotides before/after the substring for both strands.

Task:

1. Ask the user for a DNA sequence to search.
2. Check if the substring is present in the forward strand.
3. Generate the reverse complement of the DNA sequence.
4. Check if the substring is present in the reverse complement.
5. If present in the forward and/or reverse strand. Print nucleotides after the substring.

Use this DNA string for testing:

```
GCTAGTCGATCGTAGCTAGCTACGATCGTAGCATCGATCGTTAGCTAGCGTACGATCGTGCTAGCGT
```

Test searching for `ATCG` and `AAAAA`