

Bioinformatics Assignment 1: Reverse Transcription in Python

By Noah Segal-Gould

Background: We were assigned to modify the `reverseTransDNA` function in the Python program provided on Moodle 2 so that it would accept both lowercase and uppercase strings while outputting consistently uppercase strings, reverse transcribe DNA to RNA (and vice versa), return both the original nucleotide sequence which was used as the input to the function as well as its output, replace non-nucleotide characters with the letter “N” in outputs, and finally make any general changes to improve the code.

Methods: I modified the `reverseTransDNA` function to convert its input to uppercase characters, and then check to see if it contains the representative character of uracil (“U”) which would only be present in an RNA input sequence. If those characters are not present in the string, all characters representative of thymine (“T”) are replaced with one for uracil (“U”), and then through use of a dictionary all complementary nucleotides are used to replace their originals in the reversed sequence of characters. As part of a list comprehension, all characters which are not representative of the nucleotides adenine, uracil, cytosine, and guanine are replaced with a placeholder character (“N”). If the character representative of uracil (“U”) is originally present in the input string then the function performs the same steps but uracil is replaced with thymine (“T”). In both cases, the original input string as well as the result are returned as a tuple.

Results: The output of my program was as follows:

```
***** TESTS *****
CTTTCGCAATGATAC + GAAAGCGTTACTATG => CTTTCGCAATGATACGAAAGCGTTACTATG
DNA to RNA of CTTTCGCAATGATAC: ('CTTTCGCAATGATAC', 'GUAUCAUUGCGAAAG')
DNA to RNA of GAAAGCGTTACTATG: ('GAAAGCGTTACTATG', 'CAUAGUAACGCUUUC')
RNA to DNA of UAGGCCAUCCGUAAC: ('UAGGCCAUCCGUAAC', 'GTTACGGATGGCCTA')
DNA to RNA of CTTTCGCAAZZZZZZ (with errors): ('CTTTCGCAAZZZZZZ', 'NNNNNNUUGCGAAAG')
Nuclease split on CTTTCGCAATGATAC: ('CTT', 'TCGCAATGATAC')
Nuclease split on GAAAGCGTTACTATG: ('GAA', 'AGCGTTACTATG')
```

Conclusion: I successfully modified the `reverseTransDNA` function to perform the assigned tasks and my tests show that it works properly.

Acknowledgements: I worked alone on this assignment and took inspiration from the collaborative coding we did in class.