

Homework #7

Due: January 04, 2022, Wednesday, 23:59.

Lists: Bioinformatics Glimpse

DNA carries genetic instructions for the development, functioning, growth, and reproduction of all known organisms and many viruses. In this homework, you will perform basic operations on DNA codons, such as generating random DNA sequences, extracting index(es) of the desired nucleotide(s), and removing the invalid ones. Below are the detailed instructions for the functions you should write.

1. *generate_random_codons* (*n*, *nuc_list*)

This function requires two parameters:

n: The number of codons to randomly generate

nuc_list: A nucleotide list that contains characters of nucleotides desired, such as ['A','T','C','G']

This function will randomly select the characters in the *nuc_list* to construct DNA codons (length=3) *n* times, store them in the list, and return the list.

2. *remove_invalid* (*codon_list*, *invalid_codons*)

This function requires two parameters:

codon_list: A list of codons that the invalid codons will be removed from

invalid_codons: This is a list which includes invalid codons. While writing your program, use the following invalid codon list:

```
invalid_codons = ['GGA','GGT','GGG','GGC','GCA','GCT','GCC','GCG']
```

The function first asks to the user if (s)he wants to remove the invalid codons. If the answer's first letter is "y" or "Y", the function removes the invalid codons from the list (if there is any). If there is no invalid codon in the list, print a message accordingly. The function also shows how many codons are removed.

Then, the updated list must be returned by the function.

3. *extract_seq_index* (*codon_list*, *desired_seq*)

This function requires two parameters:

codon_list: A list of codons that a desired codon's index is extracted from

desired_seq: Nucleotide(s) that will be searched in the codon list

This function takes a list of codons and returns the list of index(es) of the desired nucleotide(s).

4. *main ()*

- Firstly, nucleotide list must be defined (['A','T','C','G']).
- Get from the user an integer input: the number of random codons.
- Generate the random codon list by calling *generate_random_codons* function.
- Ask if the user wants to remove the invalid codons. If invalid codons exist and the user wants to remove them, they are removed from randomly created list. And the number of codons removed is printed.
- Then, the desired nucleotide sequence is asked to the user and if it is found in the list, the function prints the list of index(es). Here you need to use updated list (the list after you remove the invalid ones).

Sample Outputs (to get the same outputs, use **random.seed(199)** in your code):

```
How many random codons do you want to generate?: 25

Codon list = ['TGT', 'TAG', 'TTC', 'TGA', 'GGC', 'GCT', 'GTG', 'TAG', 'CTT', 'TAA',
'ACC', 'GAA', 'AGA', 'ATC', 'CAG', 'ACA', 'GTC', 'GGG', 'GAG', 'AAG', 'TTT', 'CTA',
'TTC', 'GCT', 'TGC']

Do you want to remove invalid codons? (If there is any): yes

3 invalid codons are removed.

Current codon list = ['TGT', 'TAG', 'TTC', 'TGA', 'GTG', 'TAG', 'CTT', 'TAA', 'ACC',
'GAA', 'AGA', 'ATC', 'CAG', 'ACA', 'GTC', 'GAG', 'AAG', 'TTT', 'CTA', 'TTC', 'GCT',
'TGC']

Which nucleotide(s) do you want to search: AAG

Index(es) of the nucleotide(s) you searched = [16]
```

```
How many random codons do you want to generate?: 20

Codon list = ['TGT', 'TAG', 'TTC', 'TGA', 'GGC', 'GCT', 'GTG', 'TAG', 'CTT', 'TAA',
'ACC', 'GAA', 'AGA', 'ATC', 'CAG', 'ACA', 'GTC', 'GGG', 'GAG', 'AAG']

Do you want to remove invalid codons? (If there is any): no

Current codon list = ['TGT', 'TAG', 'TTC', 'TGA', 'GGC', 'GCT', 'GTG', 'TAG', 'CTT',
'TAA', 'ACC', 'GAA', 'AGA', 'ATC', 'CAG', 'ACA', 'GTC', 'GGG', 'GAG', 'AAG']

Which nucleotide(s) do you want to search: CA

Index(es) of the nucleotide(s) you searched = [14, 15]
```

```
How many random codons do you want to generate?: 12
Codon list: ['TGT', 'TAG', 'TTC', 'TGA', 'GGC', 'GCT', 'GTG', 'TAG', 'CTT', 'TAA',
'ACC', 'GAA']
Do you want to remove invalid codons? (If there is any): YEs
1 invalid codons are removed
Current codon list = ['TGT', 'TAG', 'TTC', 'TGA', 'GCT', 'GTG', 'TAG', 'CTT', 'TAA',
'ACC', 'GAA']
Which nucleotide(s) do you want to search: GCTT
There is no such sequence in the list!
```

```
How many random codons do you want to generate?: 4
Codon list = ['TGT', 'TAG', 'TTC', 'TGA']
Do you want to remove invalid codons? (If there is any): yes
There is no invalid codon in the list!
Current codon list = ['TGT', 'TAG', 'TTC', 'TGA']
Which nucleotide(s) do you want to search: G
Index(es) of the codon that you searched = [0, 1, 3]
```

Save your program as **hw07yoursurname.py**. Upload your file to Blackboard at “Homework 7”.

Make sure you place comments in your program. Place your name, ID as comments at the top. **You should only use the features and commands you have learnt until this homework was given.** You are especially not allowed to use BREAK, CONTINUE, QUIT or any other statement that prematurely stops a loop or your program.

While doing all your homework assignments, remember that:

- *You should not work together,*
- *You should not give or take any files,*
- *You should not give or take help other than simple verbal hints.*