

1. EXERCISE (8.1.pl)

Look at your code from Exercise 7.5. Right now the matching of codons is greedy. This means Perl will attempt to match the last subsequence within the long sequence that looks like a stop codon. Update the code to find the first stop codon! Is there a difference? What is more biologically meaningful?

2. EXERCISE (8.2.pl)

Store this text into a scalar variable:

“blue dog red cat red dog green cat pink dog black cat white dog”

Find out the color of the 3rd dog!

3. EXERCISE (8.3.pl)

Create a substitution pattern that automatically replaces trailing “ies” in a word with a “y” (making a singular out of a plural in the English language). If “ies” doesn’t trail the word remains unchanged. Try it out with the sentence: “flies tries babies deadliest priest fiesta”. Only the first three words should be changed...

4. EXERCISE (8.4.pl)

Store the following sentence into a single scalar variable:

“4 birds, 11 cats, 6 dogs, 9 pigs, 3 cats, 20 dogs”

Change all the individual numbers of cats and dogs to 3. So in the end the string should read: “4 birds, 3 cats, 3 dogs, 9 pigs, 3 cats, 3 dogs”

5. EXERCISE (8.5.pl)

Sometimes data has leading or trailing whitespace. Write a substitution pattern that removes leading/trailing whitespace. Try it on the three strings: “ space in front”, “space in back “ and “ space everywhere “. Hint: You can do it in one single pattern or two successive patterns.

6. EXERCISE (8.6.pl)

Remember the big.fas file from day 5? Create the reverse complement of the sequence, but in lower case. Print it to the screen. Calculate the GC-content (in %) of the sequence.

7. EXERCISE (8.7.pl)

Download the file “genes.txt” from the website. It contains a list of genes, with their names, chromosome and start and stop positions. Convert the file to be tab-separated and print it to the screen. Alternatively, print it to a new file and try opening it up in Excel!

8. EXERCISE (8.8.pl)

Update 8.7.pl to only print genes that reside on chromosome 3L or 3R.