Homework 10

Akshay Kumar Varanasi(av32826)

This homework is due on Apr. 23, 2019 at 4:00pm. Please submit as a PDF file on Canvas. Before submission, please re-run all cells by clicking "Kernel" and selecting "Restart & Run All."

Problem 1 (4 pts): Write python code that can take a string of the form "https://website.com" and of the form "https://website.com/page1", extract the name of the website (indicated here by "website"), and then print it. Make sure you get just the part between "https://" and ".com".

In [5]:
```
# You will need re to solve this problem
import re

test_string1 = "https://github.com"
test_string2 = "https://twitter.com/dariyasydykova"

# Your code goes here
def website(string):
    match = re.search("https://(.*).com",string)
    print("The website name is",match.group(1))


website(test_string1)

website(test_string2)
```

```
The website name is github
The website name is twitter
```

Problem 2 (6 pts): We will work with the E. coli genome. First, we download it:

In [2]:
```
from Bio import Entrez

Entrez.email = "akshayvaranasi@utexas.edu"

# Download E. coli K12 genome:
download_handle = Entrez.efetch(db="nucleotide", id="CP009685", rettype="gb", r
etmode="text")
data = download_handle.read()
download_handle.close()

# Store data into file "Ecoli_K12.gb":
out_handle = open("Ecoli_K12.gb", "w")
out_handle.write(data)
out_handle.close()
```

Write code that loops over all features in the E. coli genome, and counts the number of tRNAs and rRNAs that are contained within it. Use **regular expressions** to find an answer.

In [3]:
```
# You will need re and SeqIO to solve this problem
import re
from Bio import SeqIO


input_handle = open("Ecoli_K12.gb", "r")
record = SeqIO.read(input_handle,"genbank")
input_handle.close()

RNA_count = 0

for feature in record.features:
    #print(feature.type)
    match_rna=re.search("(t|r)RNA",str(feature.type))
    if match_rna:
        RNA_count+=1

print("The number of tRNAs and rRNAs contained within are",RNA_count)
```

```
The number of tRNAs and rRNAs contained within are 109
```