

BioinformaticsArmory

Alejandro Cebrián del Valle

26/08/2020

Introducción

Esta sección, siendo prácticos y como dice la propia página de Rosalind, esta sección es más una sección donde poner código que se utiliza de forma más o menos rutinaria y que sirve para que otra gente lo copie y lo utilice. Existen, como siempre, otras formas de hacerlo. En este caso, es lo que buscamos hacer, y si acaso, compararlo.

En este caso, se compara una simple cuenta de nucleótidos. La forma más simple que encuentro, es simplemente crear un diccionario de cada una de las letras. Puede hacerse mediante el método de la clase **String**. En este caso, para que con unos simples cambios permitan hacerlo con k-meros que no se solapen. **Biopython**, permite crear un objeto *secuencia*, que permite usarlo. Por lo que vemos, es más rápido Biopython (quizás, tenga que ver que no tiene que recorrer toda la cadena ;) (como el método `count()`)).

```
from Bio.Seq import Seq
from time import time
### Counting Nucleotides
def countingNucleotides(sequence):
    nucleotides = {}
    for nucl in sequence:
        if nucl not in nucleotides.keys():
            nucleotides[nucl] = 1
        else:
            nucleotides[nucl] += 1
    return nucleotides

### The program
sequence = "CTACATGCCGACATCCATGCCGTAGTGACTATTTTATGATAGGTGTTTCGGACATCCACCA" + \
"CAATCTGACCAGCGAGTTCGGGCTTCCTTCAACTAATCTTGAATCCATCATCAACCTCGT" + \
"GTTAATAGAAACGAATTATTTTCGTACTTACTTCACCCTAGCCCTTATGGTTTACCCGGCAC" + \
"CTGACTAAAGCTAACGCCCTACTCGCCTAGCCTGCCCATTAAACCCATTTCGATGCAGATG" + \
"CAACTATACTGTGCCTAGCGAAACTCGCCCGTGGCATGAAAAGAACATAGTCCTTACCT" + \
"TAGTGAGCGCACGACACGATCACACATCGAGGGTGGTCGGTTTGTCTAGGAGCGCGTGAC" + \
"AATAAAATTTTCTCATAGTTCTATGTTTTGCGGCACTTAGCTTTTCGCCCCCTCGGTGCGCG" + \
"AGAATTCTAGACCATGCGTAATCAAAAACCTTCTGGAAGAAGCACCTAAATAATCCAAAA" + \
"ACTAACACCTGGTACAAACTTAGCAATACACAGGCTAGACGATTTCGTAGAATGCCTATCG" + \
"GTACCGGTAGAGCAATCATGTTATTTTAGCATCTCTATAAGTCTTTTTTCGACGGCTACAT" + \
"CGTTATATATGGGTATTGTCCCATCATTACGAAAATTGCGGCGGCCTGATATATTCTATT" + \
"GGGCATATTGAGACAAAGCCATCAGAACTACGGCTTCGCATGTGTGGTCTACTATGTTGT" + \
"AGGACCTGGGATAAACATAATCCTTGCTTCGTTTTTAAGTAGCAGCCGACTCAAACCTG" + \
"CCCTATCAGCAACAGTCCCTCTGTTCCAGTTAAAGAAAGGTACAACAAGCTGATCAGTTAG" + \
"GACGCCGGAACAACTAAACGACCCGCGGTAATACCGGCGAAATTGGGGGAGGCGTCA" + \
"ATGTTCTAAAGGGACTGAGATATAGAAACGACTAACCAGTAGGTTACACGACTCACGTG" + \
```

```

        "TTATGCTTAGTTCTAGCGACT"
time1 = time()
print(" ".join([str(i) for i in countingNucleotides(sequence).values()])))

## 243 258 277 203

time2 = time()
##### The biopythonic way
bioSequence = Seq(sequence)
nucleotideList = [bioSequence.count("A"),
                  bioSequence.count("C"),
                  bioSequence.count("G"),
                  bioSequence.count("T")]
print(" ".join([str(x) for x in nucleotideList]))

## 277 243 203 258

time3 = time()
print("The difference between the two times:\n" + \
      str(time2 - time1) + " (The normal way)\n" + \
      str(time3 - time2) + " (The biopythonic way)")

## The difference between the two times:
## 0.010629415512084961 (The normal way)
## 0.019974946975708008 (The biopythonic way)

```