

Esame 13/02/2020

Andrea Passerini
andrea.passerini@unitn.it

Informatica

Programma python

Scrivere un programma che prenda in ingresso un file di proteine ciascuna rappresentata da :

- nome (stile fasta)
- sequenza di residui
- sequenza di etichette di legame con RNA dei residui ('+' = lega , '-' non lega)

estragga per ogni proteina la sua sottosequenza di legame più lunga, e stampi l'elenco di proteine e loro rispettiva sottosequenza ordinato per lunghezza della sottosequenza.

Esempio file di input

```
$ cat rbp_binding
>3BSU:A
HMFYAVRRGRKTGVFLTWNECRAQVDRFPAARFKKFATEDEAWAFVRK
-----+--+++++-----
.....
>1DI2:A
MPVGSLQELAVQKGWRLPEYTVAQESGPPHKREFTITCRVETFV...
-----++-----...
.....
```

Esempio esecuzione

```
> python rbp_binding_stats.py
inserire nome file: rbp_binding
1M90:M   TSKKKRQRGSRTHGGGSHKNRRGAGHRGGRGDAGRDKHEFHN
1M90:D   RQGWRRRIGNLGPWNPSRVRSTVPQQGQ
1M90:N   RKGSSRRTRFNKGRRSKRMMVNR
1M90:2   TGAGTPSQGKKNTTHTKCRRCG
1M90:C   KARGTKWPNVRGVAMNAVDH
. . . .
```

Programma python: suggerimento

Si possono implementare 5 funzioni separate:

- 1 una che legga il file dati e restituisca una mappa proteina → sequenza proteica e sequenza di etichette di legame.
- 2 una che data una mappa proteina → sequenza proteica e sequenza di etichette ne restituisca una proteina → sottosequenza di binding più lunga in essa contenuta (usando una funzione ausiliaria, vedi sotto)
- 3 una funzione ausiliaria che data una coppia sequenza proteica, sequenza di etichette di legame restituisca la sottosequenza di binding più lunga
- 4 una che data una mappa proteina → sottosequenza di binding stampi la lista di coppie proteina, sottosequenza ordinate per lunghezza della sottosequenza
- 5 una (o un main) che realizzi il programma richiesto usando le funzioni di cui sopra

Esercizi da linea di comando

Calcolare quante sequenze nella directory `fastas` contengono al massimo tre volte il motivo con acido aspartico (D) seguito da acido glutammico (E) seguiti da due aminoacidi qualunque.

Risultato atteso

50

Esercizi da linea di comando

Calcolare il numero massimo di residui di legame in una proteina del file `rbp_binding`

Risultato atteso

146

Modalita' di esecuzione e consegna

- 1 Accendere la macchina ed aspettare (la modalità esame si avvia automaticamente)
- 2 Aprire un browser e collegarsi al sito
`examina.icts.unitn.it`
- 3 Autenticarsi con le credenziali di ateneo
- 4 Scaricare il testo del compito
- 5 Realizzare il programma python come file `programma.py` e scrivere gli esercizi da linea di comando in un file di testo `linea_di_comando.txt`
- 6 Creare una cartella con *nome_cognome* e metterci i due file realizzati, comprimerla e caricarla su
`examina.icts.unitn.it`
- 7 Spegnere la macchina