

Esame 07/07/2021

Andrea Passerini
andrea.passerini@unitn.it

Informatica

Programma python

Scrivere un programma che:

- 1 prenda in ingresso un file FASTA di sequenze, con indicazione di organismo e localizzazione subcellulare (e.g. `>7B2_HUMAN:Secretory`)
- 2 per ogni organismo e localizzazione, stampi il numero di sequenze presenti, ordinando la stampa in maniera crescente nel numero di sequenze dell'organismo

Esempio file di ingresso

```
>7B2_HUMAN:Secretory
MVSRMVSTMLSGLLFWLASGWTPAFAYSPRTPDRVSEADIQRLLH..
>A1AG1_MUSCR:Secretory
MALHMILVMLSLLPLLEAQNPETHVNITIGEPITNETLSWLSDKWF..
...
>ZW10_HUMAN:Cytoplasm
ASFVTEVLAHSGRLEKEDLGTRISRLTRRVEEIKGEVCNMISKKY..
>ZYG_HUMAN:Cytoplasm
MAAPRPSPAISVSVSAPAFYAPQKKFGPVVAPKPKVNPFRPGDSE..
```

Esempio esecuzione

```
> python stats.py
```

```
Inserire nome file: seq.fasta
```

source		Cytoplasm		Secretory		Mitochondrion		Nucleus
ACAVI	0	1	0	0				
ACITR	0	1	0	0				
ACTEQ	0	1	0	0				
AGEOR	0	1	0	0				
...								
BOVIN	12	13	27	35				
RAT	27	30	7	30				
CAEEL	65	20	2	12				
DROME	188	47	18	59				
MOUSE	145	72	16	79				
HUMAN	636	203	107	191				

Programma python: suggerimento

Si possono implementare 4 funzioni separate:

- 1 una che legga il file di sequenze e restituisca una mappa da organismo a mappa localizzazione → lista di sequenze, più l'elenco delle localizzazioni (per l'intestazione)
- 2 una che prenda la mappa di mappe, e restituisca un elenco di organismi ordinato per numero totale di sequenze in ciascun organismo (a prescindere dalla loro localizzazione)
- 3 una che prenda la mappa di mappe, la lista di sequenze e l'elenco ordinato di organismi, e calcoli e stampi per ogni organismo e localizzazione il numero di sequenze corrispondenti
- 4 una (o un main) che usando le funzioni di cui sopra realizzi il programma richiesto e stampi i risultati.

Esercizi da linea di comando

Stampare l'elenco degli organismi nel file `seq.fasta` ordinati in ordine decrescente per numero di sequenze.

Risultato atteso

```
1137 HUMAN
 312 MOUSE
 312 DROME
  99 CAEEL
  94 RAT
  87 BOVIN
  47 CHICK
  28 XENLA
  18 CANFA
... 
```

Esercizi da linea di comando

Calcolare quante sequenze nel file `seq.fasta` cominciano per metionina (M), e non hanno glicine (G) nei primi 50 residui.

Risultato atteso

145

Modalita' di esecuzione e consegna

- ➊ Aprire un browser e collegarsi al sito
`examina.icts.unitn.it`
- ➋ Autenticarsi con le credenziali di ateneo
- ➌ Scaricare il testo del compito
 - Per esame da 6CFU:
Realizzare il programma python come file `programma.py`
e caricarlo su `examina.icts.unitn.it`
 - Per esame da 9CFU:
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`, creare uno zip con i due file e
caricarlo su `examina.icts.unitn.it`
- ➍ Spegnere la macchina