

Esame 13/02/2018

Andrea Passerini
andrea.passerini@unitn.it

Informatica

Programma python

Scrivere un programma che prenda in ingresso un file di variazioni genetiche, con relative conseguenze e stampi:

- 1 il numero medio per gene di variazioni con una certa conseguenza (missense, stop gained stop lost)
- 2 la frazione di variazioni per ciascuna delle possibili sostituzioni aminoacido → aminoacido

Esempio File di variazioni

UniProt - Universal Protein Knowledgebase
SIB Swiss Institute of Bioinformatics; Geneva, Switzerland
European Bioinformatics Institute (EBI); Hinxton, United Kingdom
Protein Information Resource (PIR); Washington DC, USA

.....

.....

VARIANT INDEX

Gene Name	AC	Variant	AA Change	Source	DB ID
5-MCDG	Q9PUA4	p.Leu6Val	rs736077079	missense variant	
5-MCDG	Q9PUA4	p.Pro41Leu	rs735492195	missense variant	
5-MCDG	Q9PUA4	p.Gln53Glu	rs316223795	missense variant	
5-MCDG	Q9PUA4	p.Ala343Val	rs737573019	missense variant	
5-MCDG	Q9PUA4	p.Pro364Leu	rs733456845	missense variant	
5-MCDG	Q9PUA4	p.Gln391Glu	rs15273172	missense variant	
A1CF	A0A1D5P632	p.Val163Leu	rs732446429	missense variant	
A1CF	A0A1D5P632	p.Glu200Gly	rs733574983	missense variant	
A1CF	A0A1D5P632	p.Val201Gly	rs736852055	missense variant	

.....

Esempio esecuzione

```
> python variations_stats.py
Variation file name: gallus_gallus_variation.txt
missense variant      34.34276
stop gained           0.41343
stop lost              0.07067
Leu-Val 0.00822
Pro-Leu 0.00710
Gln-Glu 0.00233
Ala-Val 0.01278
Val-Leu 0.00731
Glu-Gly 0.05398
Val-Gly 0.09690
Thr-Pro 0.07569
Asp-Ala 0.01836
Phe-Ser 0.00517
Ser-Asn 0.00325
Thr-Met 0.00436
Arg-Ser 0.00193
...
```

Programma python: suggerimento

Si possono implementare 5 funzioni separate:

- 1 una che legga il file di variazioni e restituisca un elenco di tuple con nome del gene, variazione e conseguenza.
- 2 una che prenda l'elenco letto e restituisca un dizionario con per ogni conseguenza il numero medio di volte in cui compare in un gene
- 3 una che prenda l'elenco letto e restituisca un dizionario con per ogni variazione aminoacido → aminoacido la frazione di volte in cui tale variazione compare sul totale delle variazioni
- 4 una che prenda un dizionario in ingresso e ne stampi una versione formattata
- 5 una (o un main) che realizzi il programma richiesto usando le funzioni di cui sopra

Esercizi da linea di comando

Elencare i file all'interno della directory `fastas/` aventi lunghezza (numero di caratteri) pari.

Risultato atteso

```
374 16vpA.f
108 1a5nA.f
166 1aa7A.f
364 1adt_.f
208 1ast_.f
etc..
```

Esercizi da linea di comando

Stampare a schermo i nomi dei file e i primi 5 elementi delle sequenze che:

- Hanno una (A) in quinta posizione
- Hanno una (D) in quinta posizione, ma non cominciano con (S)

Risultato atteso

```
1a3c_.f:MNQKA  
1ezvB.f:LTVSA  
1gugA.f:MSISA  
1h7dA.f:MVAAA  
1p5uB.f:ADLTA  
1rh4_.f:XAALA  
1vig_.f:INRMD
```

Modalita' di esecuzione e consegna

- 1 Avviare la macchina in modalita' `ESAME`
- 2 Autenticarsi con nome utente `sci-esame` e password fornita dal docente
- 3 Il testo del compito ed i file necessari si trovano in una cartella `Testo` sul Desktop
- 4 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`
- 5 Creare sul Desktop una cartella con *nome_cognome* e metterci i due file realizzati.
- 6 Eseguire il logout ma NON spegnere la macchina