

# Esame 04/02/2019

Andrea Passerini  
passerini@disi.unitn.it

Informatica

## Programma python

Scrivere un programma che:

- prenda un ingresso un file di allineamenti multipli
- stampi un sequenza “consenso”, contenente in ogni posizione l'aminoacido più frequente nel profilo, saltando tutte le posizioni in cui la cosa più frequente non è un aminoacido

## Esempio file allineamento

```
# STOCKHOLM 1.0
#=GF ID      RRM_1
#=GF AC      PF00076.17
#=GF DE      RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
#=GF PI      rrm;
#=GF AU      Eddy SR, Birney E
#=GF SE      Published_alignment
[...]
OCEN5_PARTE/256-312      .....kemr-----
A0CA51_PARTE/258-315    .....ekemr-----
A0C441_PARTE/141-196    .....rt-----
A0BF98_PARTE/253-307    .....ra-----
Q23D16_TETTH/112-176    .....IYVKN
D3BVL8_POLPA/132-200    .....IFIGN
```

## Esempio esecuzione

```
>>> python consensus.py
```

```
Inserire nome file: alignments
```

```
LFVGNLPPDVTEEDLRELF SKFGEVESVRVVRDTGRSKGFGFVEFEDPEDA
```

## Programma python: suggerimento

Si possono implementare 5 funzioni separate:

- 1 una che legga il file e restituisca una lista di allineamenti
- 2 una che prenda in ingresso la lista di allineamenti, per ogni posizione estragga il profilo (la colonna dell'allineamento), calcoli il carattere più frequente e lo aggiunga alla sequenza consenso se diverso da "-" e ".".
- 3 una funzione ausiliaria che dato un profilo calcoli il dizionario dei conteggi
- 4 una funzione ausiliaria che dato il dizionario dei conteggi estragga il carattere più frequente
- 5 un main che realizzi il programma richiesto usando le funzioni di cui sopra

## Shell: esercizio #1

Dato il file `utr.txt`, contenente ad ogni riga un id `ENSGXXXXXXXXXXXX` e una sequenza di rna separati da `\t`, restituire il massimo tra le lunghezze delle sequenze.

## Soluzione

8230 (8229, piú il newline contato da `wc`)

## Shell: esercizio #2

Dato il file `all.fasta`, stampare in ordine lessicografico le proteine appartenenti alla localizzazione `nuclear` — indicata nell'header delle varie sequenze.

## Esempio

La proteina citoplasmatica Q07815 è scritta:

```
>Q07815 cytoplasmic
MDGSGEQPRGGVSSRIEQGEWGGRHPSWPWTRCLMRPPRS
```

## Soluzione

```
AEETA...VESTA
...
VYPTR...VVRII
```

## 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 `utility.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