

Laboratorio di Calcolo per Fisici, Prima esercitazione valutata

Canale A-C, Docente: Nicoletta Gnan

1. Per svolgere l'esercitazione avrete 3 ore.
2. sono concessi libri di testo e appunti ed è ammesso discutere la soluzione con il proprio compagno di gruppo (a bassa voce), ma non con gli altri gruppi.
3. **L'uso di cellulari/tablet/laptop non è ammesso, pena l'annullamento del compito.**
4. Il programma va scritto e salvato esclusivamente sul computer del laboratorio, utilizzando come username **studente** e come password **informatica**.
5. Tutti i file vanno salvati in una cartella di nome **LCNG_XX all'interno** della cartella home (**/home/studente**), dove **XX** è il numero del gruppo di appartenenza. Per fare un esempio, il gruppo 100 dovrà utilizzare la cartella **/home/studente/LCNG_100**. **Tutto ciò che non si trova all'interno della cartella non verrà valutato.**
6. Per sicurezza inserite nelle prime righe del file .c tre righe di commento contenenti il nome del gruppo ed il nome, cognome e numero di matricola dei componenti del gruppo.

► **Esercizio:** Il DNA è una catena costituita da nucleotidi di tipo A, T, C, G il cui ordine costituisce l'informazione genetica. Spesso i genetisti sono interessati a piccole sequenze di nucleotidi (geni) e per questo hanno bisogno solo di piccoli frammenti di DNA. Essi utilizzano quindi degli enzimi particolari che agiscono come delle forbici molecolari che tagliano il DNA in punti specifici. Si formano così dei frammenti di DNA di lunghezza differente. Per poter continuare a lavorare con il frammento di DNA sul quale si trova il gene d'interesse, il genetista deve isolarlo. A questo proposito deve riuscire a separare un frammento di DNA avente determinata lunghezza da tutti gli altri a cui non è interessato. La separazione basata sulla lunghezza dei frammenti di DNA avviene comunemente tramite un metodo chiamato "elettroforesi su gel". **Scopo di questa esercitazione è quella di simulare l'azione di un enzima che effettua un taglio nella sequenza di DNA ogni volta che trova il nucleotide G e di caratterizzare i frammenti di DNA risultanti in base alla loro lunghezza, calcolando la distribuzione $P(l)$ della lunghezza dei frammenti .**



A tal fine si richiede di

1. produrre un codice chiamato `enzima.c` che

- definisca, attraverso una direttiva al preprocessore, la lunghezza $L_{\max} = 10000$ della sequenza iniziale di DNA;
- stampi un messaggio iniziale per spiegare all'utente che cosa fa il programma;
- chieda in input il numero di tagli N che si vuole simulare specificando che il numero deve essere compreso tra 2000 e 5000. Se questa richiesta non viene soddisfatta, il programma stampa un messaggio di errore e richiede di nuovo in input il dato;
- definisca una funzione `DesignDNA` che inizializzi una sequenza di DNA nel seguente modo:
 - La funzione inizialmente riempie un array di tipo `char` chiamato `dna` (passato in input alla funzione), con nucleotidi di tipo A, C e T. Il tipo di nucleotide da assegnare ad ogni elemento dell'array (lettera 'A', 'C' o 'T') viene scelto in maniera casuale con uguale probabilità.
 - Successivamente la funzione modifica N nucleotidi della sequenza appena generata con dei nucleotidi di tipo G. Questo viene fatto estraendo N volte in maniera casuale le posizioni dell'array dove inserire la lettera 'G'. Durante questa procedura la funzione verifica se nella posizione appena estratta sia già stato sostituito un nucleotide G e in caso estrae una nuova posizione. Inoltre gli elementi dell'array dove è possibile la sostituzione dei nucleotidi G sono tutti tranne il primo e l'ultimo elemento.
- definisca una funzione chiamata `Cut` che scorra l'array `dna` precedentemente riempito con le modalità descritte al punto sopra, e memorizzi in un array di interi chiamato `taglio` (anch'esso passato in input alla funzione) i punti della sequenza dove l'enzima effettua un taglio (ossia le posizioni dove si trovano i nucleotidi di tipo G). Ad esempio, se i primi tre nucleotide di tipo G si trovano nelle posizioni 3, 9 e 14 dell'array `dna`, allora `taglio[0]=3`, `taglio[1]=9` e `taglio[2]=14`.
- definisca una funzione chiamata `AnalisiL` che calcoli l'istogramma delle lunghezze dei frammenti della sequenza di DNA, scrivendo il risultato nel file `Histo.dat`. La lunghezza dei segmenti si può facilmente calcolare una volta noti i punti della sequenza dove sono stati effettuati i tagli (e che sono stati memorizzati nell'array `taglio`). Infatti la differenza tra le posizioni di due tagli consecutivi i e $i - 1$ (ad esempio `taglio[1]-taglio[0]=6`) corrisponde alla lunghezza del frammento creatosi da quei due tagli (un frammento di 6 nucleotidi in questo caso). Si faccia attenzione a considerare il frammento iniziale e finale della sequenza. Per maggior chiarezza si osservi la figura nel compito.

2. preparare uno script chiamato `Histo.py` che grafichi l'istogramma della distribuzione dei frammenti di DNA contenuta in `Histo.dat`. Si salvi il grafico in `Histo.png`.