

Εργαστηριακές Ασκήσεις Βιοπληροφορικής

Άσκηση 9η “C, μέρος τέταρτο”

Όνοματεπώνυμο :

Αριθμ. Μητρώου :

Ύλη ασκήσεως :

- Συγγραφή, μεταγλώττιση και εκτέλεση δύο μικρών προγραμμάτων.
-

Δημιουργήστε τον υποκατάλογο `~/practicals/9/` και κάντε τον τον τρέχοντα κατάλογο σας.

Αντιγράψτε από τον κατάλογο `/c0d4/people/biologic/nglykos/pract/` το αρχείο με το όνομα `standard.dat`, στον τρέχοντα κατάλογο σας (δηλ. στον κατάλογο `~/practicals/9/`). Όπως μπορείτε να δείτε μέσω του κειμενογράφου αρεσκείας σας, το αρχείο αυτό περιέχει 100 αριθμούς κινητής υποδιαστολής (έναν ανά γραμμή του αρχείου).

Γράψτε ένα πρόγραμμα σε C (με το όνομα `mean.c`) το οποίο για όσους αριθμούς x_i (με $1 \leq i \leq N$) του δωθούν (από την καθιερωμένη είσοδο) να υπολογίζει και να τυπώνει (στην καθιερωμένη έξοδο) το μέσο όρο (μ) και την τυπική απόκλιση (σ) τους. Αυτά τα μεγέθη ορίζονται ως :

$$\mu = \sum_{i=1}^N \frac{x_i}{N}$$

$$\sigma = \sqrt{\sum_{i=1}^N \frac{(x_i - \mu)^2}{N - 1}}$$

Μεταγλωττίστε το πρόγραμμα που δημιουργήσατε και μετονομάστε το προκύπτον εκτελέσιμο αρχείο σε `mean` (από `a.out` που ήταν, βέβαια, το αρχικό του όνομα). Για τη μεταγλώττιση θα πρέπει να χρησιμοποιήσετε μια εντολή του τύπου `cc mean.c -lm`

Χρησιμοποιήστε το πρόγραμμα σας (το `mean`) για να υπολογίσετε το μέσο όρο και την τυπική απόκλιση των αριθμών που περιέχονται στο αρχείο `standard.dat`. Ποιές τιμές βρήκατε ;

Αντιγράψτε από τον κατάλογο /c0d4/people/biologic/nglykos/pract/ το αρχείο με το όνομα DNA.seq, στον τρέχοντα κατάλογο σας (δηλ. στον κατάλογο ~/practicals/9/). Όπως μπορείτε να δείτε μέσω του κειμενογράφου αρεσκείας σας, το αρχείο αυτό περιέχει μια αλληλουχία DNA (η οποία είναι η κωδικοποιούσα αλληλουχία DNA για την πρωτεΐνη βακτηριοροδοψίνη από το οργανισμό *Halobacterium salinarum*).

Γράψτε ένα πρόγραμμα σε C (με το όνομα composition.c) το οποίο να διαβάζει από την καθιερωμένη είσοδο μια αλληλουχία DNA και να υπολογίζει και να τυπώνει (στην καθιερωμένη έξοδο) τα κάτωθι :

1. Τον συνολικό αριθμό από κάθε ένα από τα νουκλεοτίδια της αλληλουχίας που διαβάστηκε από την είσοδο (πόσα A, πόσα T, ...).
2. Την επί τοις εκατό περιεκτικότητα της αλληλουχίας για κάθε ένα από τα νουκλεοτίδια (επί τοις εκατό ποσοστό του A, επί τοις εκατό ποσοστό του T, ...).
3. Τον συνολικό αριθμό και επί τοις εκατό περιεκτικότητα της αλληλουχίας σε πούρινο-νουκλεοτίδια (A και G), πυριμιδινό-νουκλεοτίδια (T και C) και ζευγών GC (G και C).

Μεταγλωττίστε το πρόγραμμα που δημιουργήσατε και μετονομάστε το προκύπτον εκτελέσιμο αρχείο σε composition (από a.out που ήταν, βέβαια, το αρχικό του όνομα).

Χρησιμοποιήστε το πρόγραμμα σας (το composition) για να υπολογίσετε τα στατιστικά στοιχεία που αναφέρθηκαν προηγουμένως για την αλληλουχία που περιέχεται στο αρχείο DNA.seq. Ποιές τιμές βρήκατε ;

Ο έλεγχος των προγραμμάτων σας θα γίνει κατευθείαν από τους καταλόγους ~/practicals/9/ των περιοχών σας. Αυτοί θα πρέπει να περιέχουν όλα τα αρχεία της άσκησης, δηλαδή τα : standard.dat DNA.seq mean.c mean composition.c composition
Σημειώστε ότι :

1. Τα permissions τόσο των καταλόγων όσο και των αρχείων που περιέχονται σε αυτούς, θα πρέπει να είναι τέτοια ώστε ένας χρήστης που ανήκει στους κατηγορία "other" να μπορεί να διαβάζει (και στην περίπτωση των εκτελέσιμων αρχείων, να εκτελεί) τα αρχεία σας και καταλόγους σας (πάντα για το ~/practicals/ ...).
2. Τα αρχεία αυτής της άσκησης πρέπει να υπάρχουν (με τα σωστά permissions) στις περιοχές όλων ανεξαιρέτως (εάν εργαζόσασταν ανά δύο, θα πρέπει ο δεύτερος να αντιγράψει όλα τα αρχεία στην περιοχή του και να τους δώσει τα σωστά permissions).

ΕΑΝ Ο ΕΛΕΓΧΟΣ ΤΩΝ ΑΡΧΕΙΩΝ ΤΗΣ ΑΣΚΗΣΗΣ ΚΑΤΑΣΤΕΙ ΑΔΥΝΑΤΟΣ, Ο ΙΔΙΟΚΤΗΤΗΣ ΤΟΥΣ ΘΑ ΒΑΘΜΟΛΟΓΗΘΕΙ ΜΕ ΒΑΘΜΟ ΙΣΟ ΜΕ ΤΟ ΧΑΜΗΛΟΤΕΡΟ ΒΑΘΜΟ ΠΟΥ ΕΧΕΙ ΔΩΘΕΙ ΣΤΟ ΠΑΡΕΛΘΟΝ ΓΙΑ ΤΑ ΕΡΓΑΣΤΗΡΙΑ ΑΥΤΟΥ ΤΟΥ ΜΑΘΗΜΑΤΟΣ.