**ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ-BIONFORMATICS**

ΑΛΕΞΑΝΔΡΟΣ ΣΚΑΡΠΕΛΟΣ (Π17122) – ΔΗΜΗΤΡΙΟΣ ΜΑΤΣΑΓΓΑΝΗΣ (Π17068)

Περιεχόμενα

[**ΑΣΚΗΣΗ 11.4** 3](#_Toc43476985)

[1.ΕΚΦΩΝΗΣΗ ΕΡΓΑΣΙΑΣ 3](#_Toc43476986)

[2.ΑΠΑΡΑΙΤΗΤΑ ΕΡΓΑΛΕΙΑ 4](#_Toc43476987)

[3.ΥΛΟΠΟΙΗΣΗ 5](#_Toc43476988)

[3.1.ΔΗΛΩΣΕΙΣ ΜΕΤΑΒΛΗΤΩΝ 5](#_Toc43476989)

[4.ΑΠΟΤΕΛΕΣΜΑΤΑ 6](#_Toc43476990)

# **ΑΣΚΗΣΗ 11.4**

## 1.ΕΚΦΩΝΗΣΗ ΕΡΓΑΣΙΑΣ

Στο σχήμα 11.7 φαίνεται ένα HMM με δύο καταστάσεις α και β. Όταν το HMM βρίσκεται στην κατάσταση α, έχει την μεγαλύτερη πιθανότητα να εκπέμψει πουρίνες (Aκαι G). Όταν βρίσκεται στην κατάσταση β, έχει μεγαλύτερη πιθανότητα να εκπέμψει πυριμιδίνες (Cκαι T). Αποκωδικοποιήστε την πιο πιθανή ακολουθία των καταστάσεων (α/β) για την αλληλουχία GGCT. Χρησιμοποιήστε λογαριθμικές βαθμολογίες για κανονικές βαθμολογίες πιθανοτήτων.

## 2.ΑΠΑΡΑΙΤΗΤΑ ΕΡΓΑΛΕΙΑ

Η εργασία μας υλοποιήθηκε στο προγραμματιστικό-εκπαιδευτικό εργαλείο matlab το οποίο πρέπει να είναι εγκατεστημένο. Απαραίτητο εργαλείο-βιβλιοθήκη αποτελεί το bionformantics tool που πρέπει να είναι και αυτό εγκατεστημένο για την υλοποίηση της. Κάνουμε ιδιαίτερη αναφορά σε αυτήν την βιβλιοθήκη επειδή ο χρήστης θα πρέπει να φροντίσει να την επιλέξει κατά την εγκατάσταση του matlab μέσω installer.

## 3.ΥΛΟΠΟΙΗΣΗ

Το πρόγραμμα μας αποτελείται από 1 αρχείο το οποίο πρέπει και να εκτελέσουμε ώστε να δούμε τα αποτελέσματα. Σε αυτήν την άσκηση προσπαθούμε να ανακαλύψουμε την ‘καλύτερη’ διαδρομή (αυτή με το μεγαλύτερο σκορ σαν σύνολο) για την κωδικοποίηση του GGCT.

## 3.1.ΔΗΛΩΣΕΙΣ ΜΕΤΑΒΛΗΤΩΝ

Θεωρούμε ισοπίθανες αρχικές πιθανότητες να ξεκινάει είτε από την κατάσταση Α είτε από την B (0.5 για κάθε μια) εφόσον το πρόγραμμα δεν μας υποδεικνύει κάτι διαφορετικό. Έπειτα δημιουργήσαμε έναν πίνακα που αναφέρεται στην πιθανότητα να πηγαίνει από την μία κατάσταση σε άλλη ή ίδια να πηγαίνει στον εαυτό της. Η ακολουθία GGCT μετατρέπετε σε αριθμούς μέσω της εντολής nt2int για να μπορεί να γίνει αντιληπτή. Επομένως μετατρέπεται σε ακολουθία αριθμών (3,3,2,4).

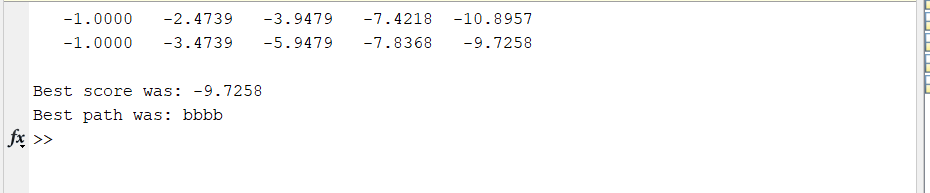
## 3.2.ΚΥΡΙΟ ΜΕΡΟΣ

Αρχικά, δημιουργούμε έναν πίνακα δύο διαστάσεων ο οποίος αποτελείται μόνο από μηδενικά. Έπειτα προσπαθούμε μέσω εμφολευμένων loops να ακολουθήσουμε όλες τις πιθανές διαδρομές που προκύπτουν από το διάγραμμα που κατασκευάζουμε στο χαρτί. Πιο συγκεκριμένα, κοιτάμε κάθε φορά ποιο στοιχείο εξετάζουμε (πρώτο loop) και στην συνέχεια ελέγχουμε σε ποια κατάσταση βρισκόματσε και έπειτα τα αποθηκεύουμε στον πίνακα που υπήρχαν τα μηδενικά ο οποίος χρησιμεύει σαν ένας λευκός πίνακας πάνω στον οποίο γράφουμε τις αντλούμενες πληροφορίες.

## 3.3.ΤΕΛΙΚΟ ΣΤΑΔΙΟ

Τέλος υπολογίζεται το συνολικό σκορ των δύο διαδρομών(στο συγκεκριμένο πρόβλημα, σε κάποιο άλλο θα μπορούσε να έχει παραπάνω καταστάσεις άρα και παραπάνω διαδρομές), συγκρίνονται και αποφασίζεται ποια διαδρομή είναι η καλύτερη.

## 4.ΑΠΟΤΕΛΕΣΜΑΤΑ



Όπως φαίνεται και στην παραπάνω φωτογραφία συγκρίνονται στο τέλος οι τιμές του πίνακα (το 1ο και το 2ο στοιχείο της 5ης στήλης) και επιλέγεται η ο μεγαλύτερος ο αριθμός δηλαδή η διαδρομή με το καλύτερο σκορ. Άρα επιλέγουμε την διαδρομή ‘bbbb’ αφού είναι αυτή που μας δίνει το καλύτερο αποτέλεσμα.