



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΒΙΟΪΑΤΡΙΚΗ

## ΕΡΓΑΣΙΑ 2

# Υλόποιήθηκε για το μάθημα “ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΛΟΓΙΑ”

ΦΟΙΤΗΤΗΣ

[ΛΑΖΑΡΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ-ΠΑΝΑΓΙΩΤΗΣ, 1639]

ΚΑΘΗΓΗΤΡΙΑ

[ΑΡΤΕΜΙΣ ΧΑΤΖΗΓΕΩΡΓΙΟΥ]

ΛΑΜΙΑ, 2021-2022

## Περίληψη

Το κείμενο αυτό, αφορά στις θεωρητικές ερωτήσεις της δεύτερης εργασίας του μαθήματος "ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΛΟΓΙΑ".

### 1 Τα BED αρχεία χρησιμοποιούνται για την αποθήκευση γονιδιωμάτων περιοχών. Να αναφέρετε τι πληροφορίες αποθηκεύουμε σε κάθε μία από τις στήλες του αρχείου. Ποιες στήλες είναι υποχρεωτικές και ποιες προαιρετικές;

Το BED format δημιουργήθηκε κατά τη διάρκεια του προγράμματος του ανθρώπινου γονιδιώματος. Ένα BED αρχείο αποτελείται συνολικά από δώδεκα (12) στήλες εκ των οποίων τρεις (3) είναι υποχρεωτικές και εννέα (9) είναι προαιρετικές.

**Οι τρεις (3) υποχρεωτικές στήλες είναι οι εξής:**

1. **chrom:** Όνομα χρωμοσώματος
2. **chromStart:** Αρχική θέση μιας γονιδιακής περιοχής ενδιαφέροντος/χαρακτηριστικού σε ένα χρωμόσωμα. Η πρώτη βάση σε ένα χρωμόσωμα βρίσκεται στην θέση μηδέν (0).
3. **chromEnd:** Τελική θέση μιας γονιδιακής περιοχής/χαρακτηριστικού σε ένα χρωμόσωμα.

**Οι εννέα (9) προαιρετικές στήλες είναι οι εξής:**

1. **name:** Το όνομα της περιοχής ενδιαφέροντος/χαρακτηριστικού.
2. **score:** Σκορ το οποίο παίρνει τιμές από μηδέν (0) έως χίλια (1000).
3. **strand:** Καθορίζει την αλυσίδα του DNA που βρίσκεται η περιοχή ενδιαφέροντος/χαρακτηριστικό (χρησιμοποιούνται τα σύμβολα ".", "-", και "+").
4. **thickStart:** Η αρχική θέση στην οποία η περιοχή ενδιαφέροντος ζωγραφίζεται πιο έντονα.
5. **thickEnd:** Η τελική θέση στην οποία η περιοχή ενδιαφέροντος ζωγραφίζεται πιο έντονα.
6. **itemRgb:** Τιμή RGB της μορφής R,G,B. Αν είναι ενεργοποιημένη, η τιμή RGB καθορίζει την χρωματική απεικόνιση των δεδομένων που βρίσκονται στη συγκεκριμένη γραμμή BED.
7. **blockCount:** Ο αριθμός των εξωνίων ενός αρχείου BED.
8. **blockSizes:** Μια λίστα που είναι διαχωρισμένη με κόμματα (comma-separated). Περιλαμβάνει τα μεγέθη των εξωνίων ενός αρχείου BED
9. **blockStarts:** Μια λίστα που είναι διαχωρισμένη με κόμματα (comma-separated). Περιλαμβάνει όλες τις θέσεις εκκίνησης των εξωνίων ενός αρχείου BED. Όλες οι θέσεις πρέπει να υπολογίζονται σε σχέση με το chromStart.

Σε περίπτωση που δεν υπάρχει κάποια θέση που να ζωγραφίζεται έντονα, τα thickStart και thickEnd παίρνουν την ίδια τιμή με το chromStart.

### 2 Ο προσδιορισμός της αρχής και ο προσδιορισμός του τέλους μιας περιοχής σε ένα BED αρχείο ακολουθούν αρίθμηση που ξεκινάει από το μηδέν (zero-based) ή από το 1 (one-based);

Η αρχή μιας περιοχής ενδιαφέροντος/χαρακτηριστικού ενός BED αρχείου ελέγχεται από την στήλη chromStart ακολουθεί αρίθμηση που ξεκινάει από το μηδέν (0). Είναι δηλαδή zero-based.

Το τέλος μια περιοχής ενδιαφέροντος/χαρακτηριστικού ενός BED αρχείου ελέγχεται από την στήλη chromEnd η οποία ακολουθεί αρίθμηση που ξεκινάει από το ένα (1). Είναι δηλαδή one-based.

Η βάση που αντιστοιχεί στο chromEnd δεν περιλαμβάνεται στην απεικόνιση του γονιδίου/χαρακτηριστικού. Όμως περιλαμβάνεται ο αριθμός σε format θέσης (1-based).

Για παράδειγμα, οι πρώτες εκατό (100) βάσεις του χρωμοσώματος ένα (1), έχουν ως αρχική θέση την θέση μηδέν (0) (δηλαδή chromStart=0) και ως τελική θέση, την θέση εκατό (100) (chromEnd=100).

Στο παραπάνω παράδειγμα, οι πρώτες εκατό (100) βάσεις του χρωμοσώματος ένα (1), βρίσκονται στις θέσεις μηδέν (0) έως ενενήντα εννέα (99) (δηλαδή: 0-99) (στο one-based σύστημα οι θέσεις αυτές είναι: ένα (1) έως εκατό (100) (δηλαδή: 1-100)).

**3 Σε ένα BED αρχείο υπάρχει η παρακάτω γραμμή που προσδιορίζει μια γονιδιωματική περιοχή. Τι σημαίνει η συντομογραφία «chrM»; Σε ποια θέση του χρωμοσώματος ξεκινάει η γονιδιωματική περιοχή; Σε ποια θέση του χρωμοσώματος τελειώνει η γονιδιωματική περιοχή; Ποιο είναι το μήκος αυτής της γονιδιωματικής περιοχής;**

chrM	8366	8572	ATPase8	0	+
------	------	------	---------	---	---

Σχήμα 1: Η γραμμή για την οποία γίνεται αναφορά

1. **chrM**: Η συγκεκριμένη συντομογραφία υποδεικνύει ότι πρόκειται για μιτοχονδριακό DNA.
2. **8366**: Η αρχική θέση του μιτοχονδριακού DNA (είναι zero-based).
3. **8572**: Η τελική θέση του μιτοχονδριακού DNA (είναι one-based).

Η αρχική θέση δίνεται από την στήλη chromStart η οποία είναι zero-based. Επομένως σε one-based σύστημα η αρχική θέση είναι η 8367.

Η τελική θέση είναι η 8572.

Το μήκος της γονιδιωματικής περιοχής προκύπτει με αφαίρεση της τελικής θέσης που είναι one-based, μείον την αρχική θέση που είναι zero-based. Επομένως το μήκος της γονιδιωματικής περιοχής είναι: 206 βάσεις.

## 4 Μερos B: Προγραμματιστική εργασία

Ο κώδικας της εργασίας είναι διαθέσιμος στο github (project02):

[https://github.com/kostaslaazaros/compbio\\_projects](https://github.com/kostaslaazaros/compbio_projects)



Σχήμα 2: Qr code για το github repository