|  |
| --- |
| Photo displaying partial image of two pie charts on a canvas-textured page |
| Εργασία Βιοπληροφορικής  εαρινό εξάμηνο ακαδημαϊκού έτους 2023-2024 |
| |  |  |  | | --- | --- | --- | | ΜΑΤΙΝΑ ΠΑΠΑΔΑΚΟΥ - Π21127 |  | ΓΕΩΡΓΙΟΣ ΧΡΙΣΤΟΠΟΥΛΟΣ – Π20206 | |

ΠΕΡΙΕΧΟΜΕΝΑ

[**ΘΕΜΑ i 2**](#thema1)

[(a)2](#thema1a)

[(b)2](#thema1b)

[(c)3](#thema1c)

[**ΘΕΜΑ ii 4**](#thema2)

[**ΘΕΜΑ iii 5**](#thema3)

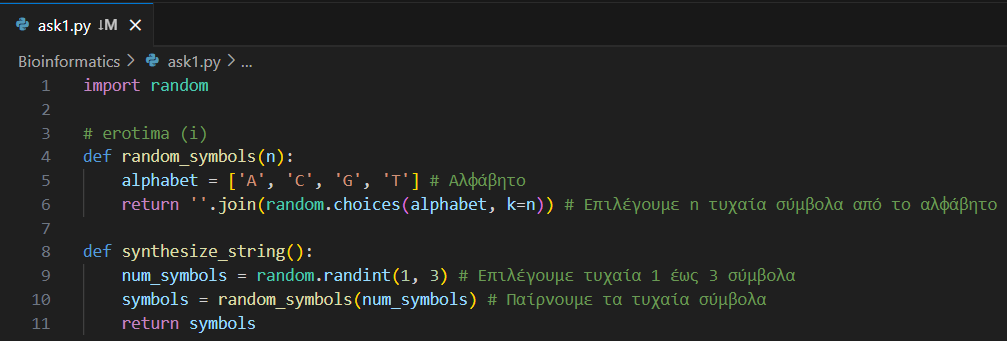
[**Απεικόνιση ενός HMM Profile με πολλαπλή στοίχιση 5**](#thema3)

**ΘΕΜΑ i**

**(a)**

Αφού καλέσουμε την βιβλιοθήκη random, θα χρειαστούμε 2 μεθόδους για το πρώτο ερώτημα.

Η synthesize\_string() επιλέγει έναν αριθμό από το ένα έως το τρία, καλεί την random\_symbols(n) και αυτή επιστρέφει ενα string με n σύμβολα που διαλέγει τυχαία από την λίστα alphabet .

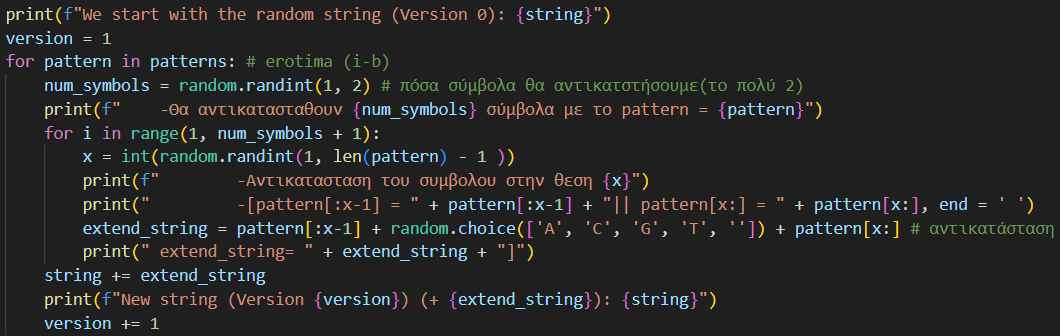


Τα τυχαία σύμβολα θα καταχωρηθούν στην συμβολοσειρά string :



**(b)**

Για κάθε ένα από τα 4 patterns θα εκτελεστεί το block κώδικα μέσα στην for. Οι εντολές print χρησιμοποιήθηκαν για την καλύτερη κατανόηση του αναγνώστη. Η extend\_string σπάει το pattern στα 3 μέρη ως εξής:  
Εστω το pattern = ΑΑΤΤGA, αν επιλέξουμε τυχαία να αντικαταστησουμε το γράμμα στην θεση 5, το G, τότε pattern[:x-1] = AATT και pattern[x:] = A και με την βοήθεια της random.choice([‘A’, ‘C’,’G’, ‘T’ , ‘ ‘]) επιλέγεται τυχαία ένα γράμμα ή το κενο . Αρα AATT + A + A = AATTAA.

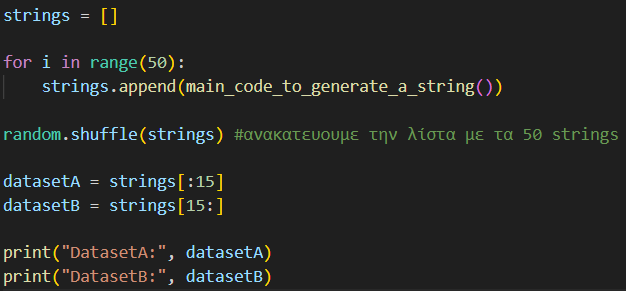
****

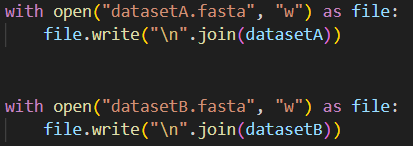
**(c)**

Τελειώνοντας το [**ΘΕΜΑ I**](#thema1), θα χρησιμοποιήσουμε πάλι την random\_symbols(n), για να μας επιστρέψει ενα string με ένα εώς δύο τυχαία σύμβολα.



Μένει να αναφέρουμε το πώς δημιουργήσαμε 50 συμβολοσειρές και τις καταχωρησαμε τυχαία σε 2 σύνολα datasetA[15] και datasetB[35]:

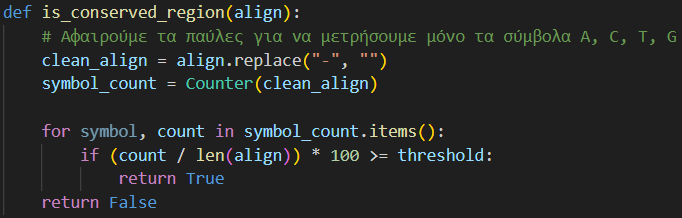


Και πώς αποθηκεύουμε στα αρχεία FullDataset, datasetA, datasetB τις συμβολοσειρες 

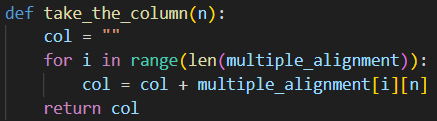
**ΘΕΜΑ ii**

**ΘΕΜΑ iii**

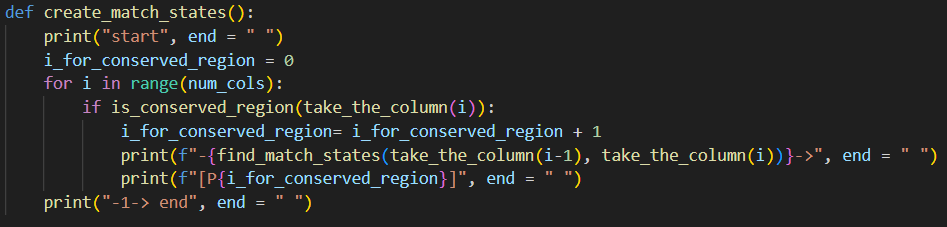
* Η πρώτη συνάρτηση που συναντάμε είναι η **is\_conserved\_region(align)**. Μετράει το πλήθος κάθε συμβόλου που υπάρχει σε κάθε στήλη της πολλαπλής στοίχισης και αν υπάρχει κάποιο σύμβολο που εμφανίζεται σε ένα συγκεκριμένο ποσοστό(π.χ. Threshold = 70%) επιστρέφει **True**, αλλιώς επιστρέφει **False**.

****

* Η συνάρτηση **take\_the\_column(n)** φτιάχνει μια συμβολοσειρά, την **col** οπου την επιστρέφει στο τέλος. Αυτή η συμβολοσειρά είναι μια στήλη του πίνακα **multiple\_alignment**, ανάλογα με το n >=0 && n <= cols. Η κύρια χρήση της, είναι να ετοιμάσει την συμβολοσειρά για να την αξιοποιήσει η παραπάνω συναρτηση και να ελέγξει αν ειναι **conserved\_region**.

****

* Η συνάρτηση **create\_match\_states()**  βρίσκει ποιες καταστάσεις είναι match.

****

Ενα παράδειγμα εκτέλεσης της **create\_match\_states()**  είναι το παρακάτω:



**Απεικόνιση ενός HMM Profile με πολλαπλή στοίχιση:**

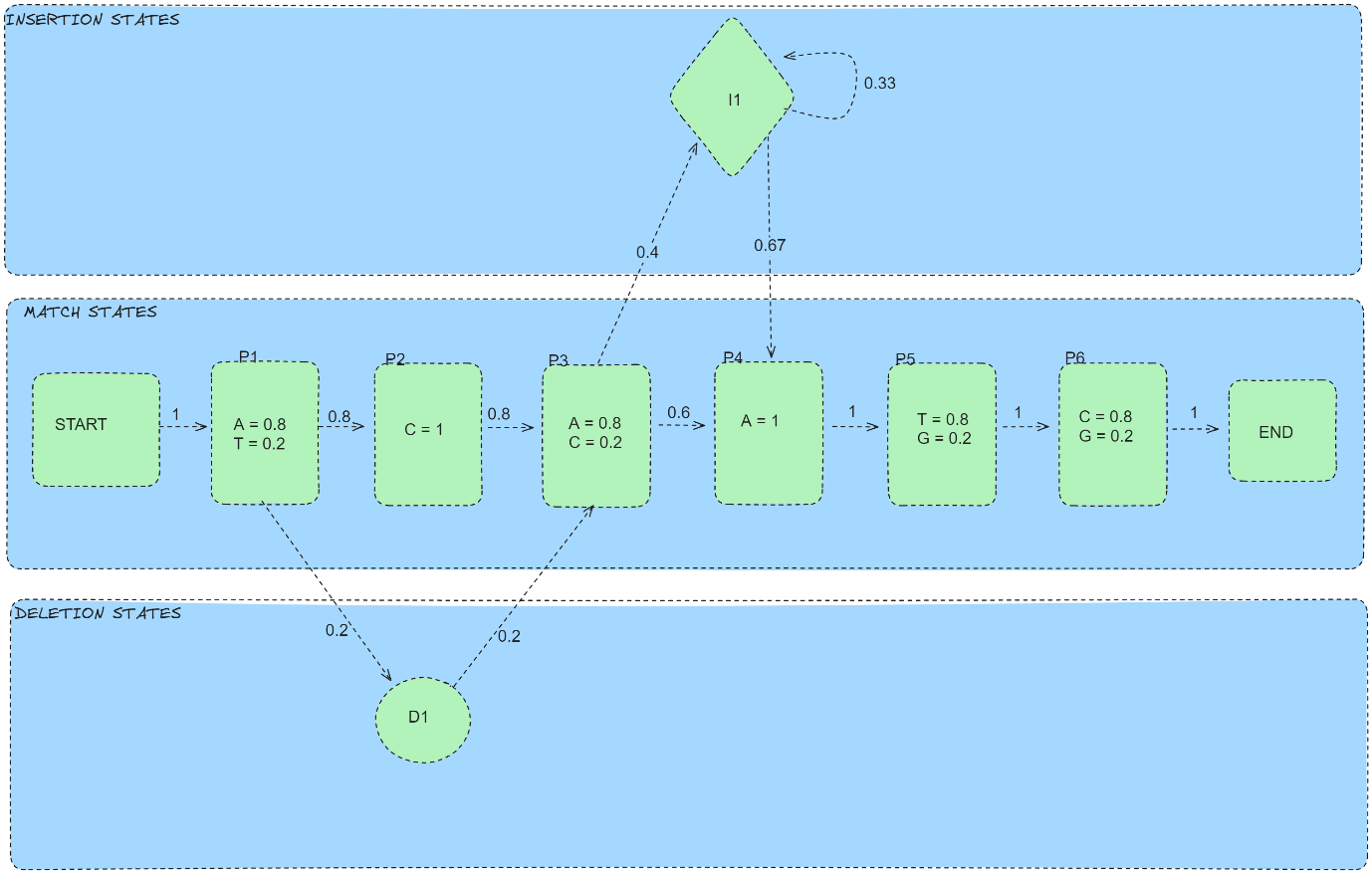
    ACA\_ \_ \_ATG

    TCAAC TATC

    ACAC\_\_AGC

    ACA\_\_ \_A TC

    A\_C\_\_ \_A TC

****