**Βιοπληροφορική 1o Project 2021-2022**

**Ομάδα 2 φοιτητών  
Τμήμα: ΗΜΤΥ**

Ονοματεπώνυμο: Κοντούλης Αναστάσιος

ΑΜ: 1066489

Έτος: 4ο

Ονοματεπώνυμο: Παπαγεωργίου Αντώνιος

ΑΜ: 1066591

Έτος: 4ο

**Ασκήσεις**

Ερώτημα 1:

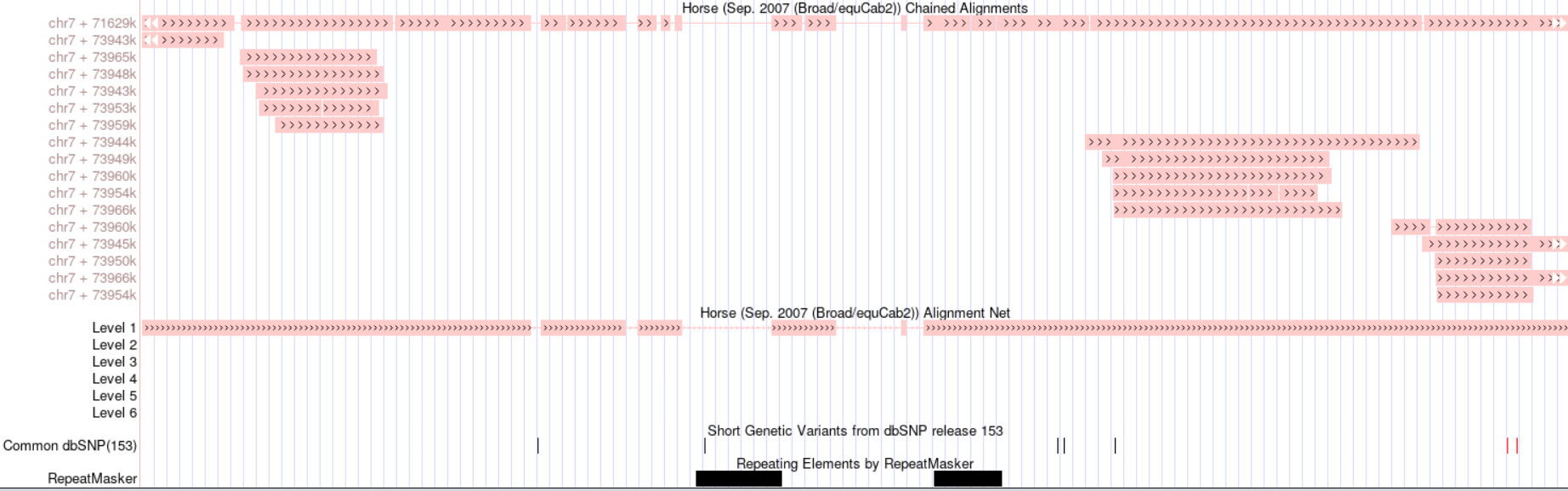
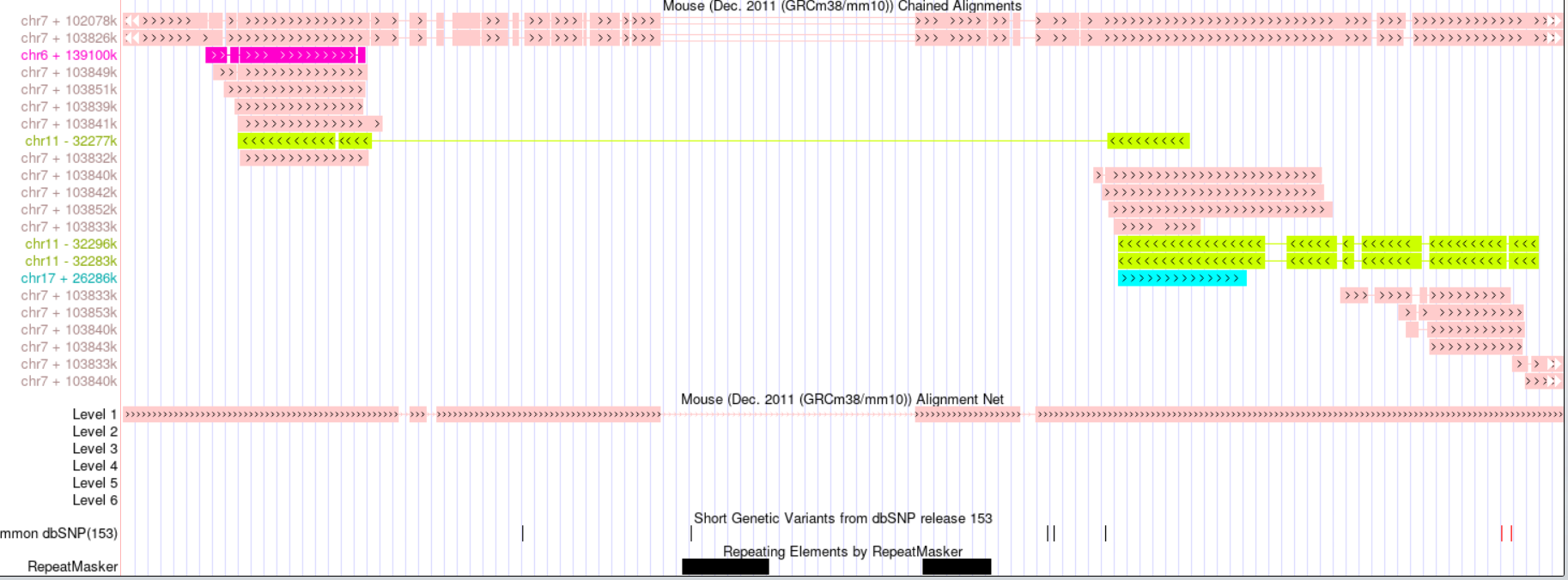
Παρακάτω θα παρουσιαστούν κάποια από τα έτοιμα εργαλεία στο Armory της Rosalind και περιληπτικά η λειτουργία τους.

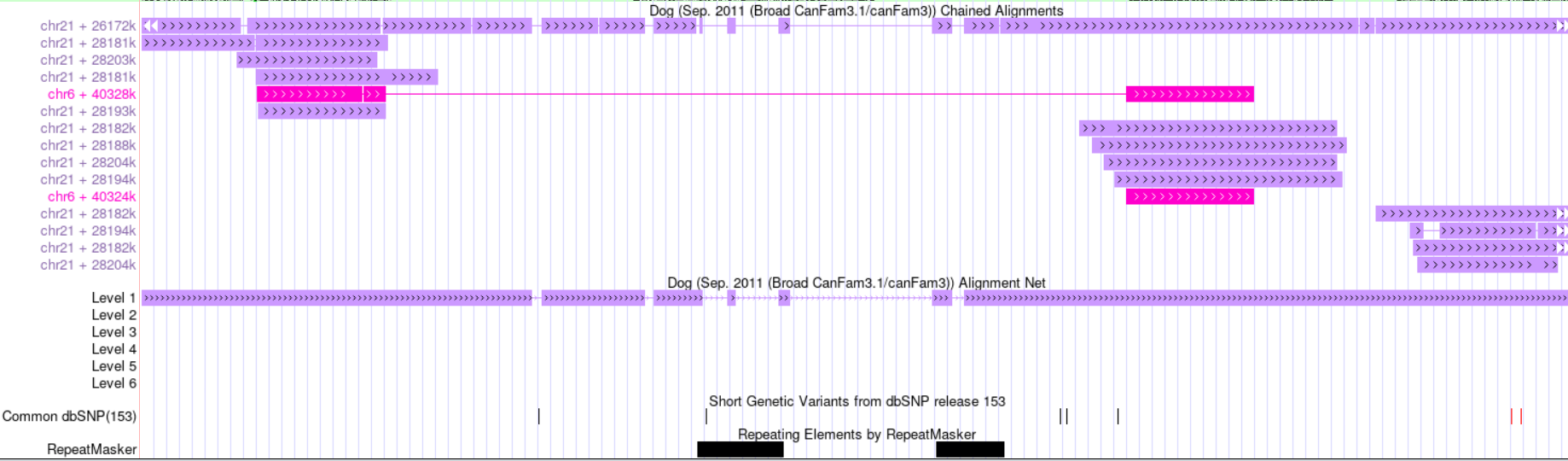
*Introduction to the Bioinformatics Armory*  
Εδώ χρησιμοποιείται το online εργαλείο DNA stats, το οποίο δέχεται ως είσοδο ένα DNA string σε απλή ή FASTA μορφή και επιστρέφει το πλήθος εμφανίσεων του κάθε συμβόλου και το ποσοστό εμφάνισης τους.  
Με τη χρήση του εργαλείου λύνεται σύντομα και απλά το αντίστοιχο πρόβλημα μέτρησης των νουκλεοτιδίων σε ένα DNA string.

*Complimenting a Strand of DNA*  
Εδώ δίνονται ως είσοδος κάποια DNA strings και ζητείται ως έξοδος ο αριθμός αυτών που ταυτίζονται με το αντίστροφα συμπληρωματικά τους, δηλαδή τα συμπληρωματικά τους γραμμένα από τα αριστερά προς τα δεξιά.  
Αυτό το πρόβλημα μπορεί να λυθεί γρήγορα χρησιμοποιώντας το εργαλείο Reverse Complement, το οποίο υπολογίζει τα reverse complements των DNA strings που παίρνει ως είσοδο, καθιστώντας το έτσι εύκολο να συγκρίνουμε με τα αρχικά.

*Read Quality Distribution*  
Εδώ παρουσιάζεται το εργαλείο FastQC το οποίο έχει online version που τρέχει στο περιβάλλον Galaxy αλλά και local version που τρέχει με Java (JRE). Ως είσοδο δέχεται μια τιμή quality threshold και FASTQ συμβολοσειρές, επιστρέφοντας τον αριθμό των δεδομένων των οποίων το μέσο quality είναι κάτω από την τιμή που δόθηκε στην είσοδο.

Ερώτημα 2:

Human-Horse  
Human-Mouse  


Human-Dog  


Γίνεται σύγκριση του γονιδίου beta-globin στα DNA strands μεταξύ ανθρώπου και διάφορων ειδών ζώων. Στα σχήματα φαίνονται τα alignments του ΗΒΒ στα επιλεγμένα είδη. Επίσης οι διαφορές προκύπτουν σε διαφορετικά χρωμοσώματα σε κάθε είδος και εμφανίζονται με χρωματικό κώδικα, στον σκύλο για παράδειγμα έχουμε τα alignments στο chr21 με μωβ χρώμα και στο chr6 με ροζ.   
  
  
Ερώτημα 3:

Αρχικά στην προεπεξεργασία κατασκευάζουμε το suffix tree του Τ χρησιμοποιώντας τον αλγόριθμο του Ukkonen σε Ο(n) χρόνο.

Το δέντρο αποθηκεύει σε κάθε κόμβο το index του συνόλου των χαρακτήρων που περιέχει.  
Επομένως θα κάνουμε αναζήτηση στο δέντρο για το string P και θα εξετάσουμε αν βρίσκεται πριν το k.

-----------------

Root = tree.root

i=0

flag = 0

**While** flag==0 **do**

**For** leaf in root.leaves **do**

Start = leaf.start

End = leaf.end

*# Αν το το node ταυτίζεται με τη συνέχεια του Ρ τότε πηγαίνουμε στον επόμενο*

**If** P[i:end-start+1] == T[start:end] **then**

*# Αν έχουμε ελέγξει όλο το Ρ*

**if** i == m-1 **then**

print(“T found in P”)

*# θέση οπού τελειώνει το Τ στο Ρ*

k\_found = end

flag = 1

**if** K\_found <= k **then**

print(“T found before k”)

**else**

print(“T found after K”)

**else**

root = leaf

*# Aν αυτός ήταν ο τελευταίος κόμβος*

**if** root.leaf == False **then**

print(“T not found in P”)

flag ==1

i = i + end - start +1

**break**

Ο αλγόριθμος προϋποθέτει ότι τα δυο αλφάβητα των Ρ, Τ έχουν τουλάχιστον ένα κοινό χαρακτήρα μεταξύ τους.

Ερώτημα 7:

Αρχικά θα αναφέρουμε πως υπολογίζεται το LCS. Είναι παρόμοιος τρόπος με τον υπολογισμό απόστασης, όμως εδώ δεν υπάρχει η δυνατότητα του replace, αρά έχουμε:

**Si,j = max( Si-1,j , Si,j-1, Si-1,j-1 +1) if vi=wj.**

Θέλουμε να δείξουμε ότι **D(n,m) = m+n -2u** :

Η μέγιστη απόσταση Dax υπάρχει όταν οι 2 συμβολοσειρές έχουν τελείως διαφορετικούς χαρακτήρες μεταξύ τους. Τότε στον πίνακα DP η οπισθοδρόμηση θα γίνεται προς τα πάνω και προς τα αριστερά, όχι διαγώνια γιατί **r=2> d=1** άρα το replace κοστίζεις περισσότερό σε κάθε κελιά - σε κάθε βλήμα η απόσταση θα αυξάνεται +1 οπότε D = n+m που είναι και οι διαστάσεις του πίνακα.

Αν υπήρχε ένα κοινό στοιχείο μεταξύ των S1 και S2 ( u=1 ) τότε σε εκείνο το σημείο η οπισθοδρόμηση θα πήγαινε διαγώνια ( e=0 ) μειώνοντας την απόστασή κατά 2 (1 αριστερά, 1 πάνω)

Οπότε θα έχουμε **D(n,m) = m+n – 2\*1**.

Ο ίδιος τύπος θα ίσχυε και για u=2 γιατί για κάθε στοιχείο το οποίο αντιστοιχίζεται στα S1,S2 σε εκείνο το κελιά οπισθοδρομούμε διαγώνια

οπότε γενικεύοντας για μήκος LCS = u : **D(n,m) = m+n -2u**.

Ο αλγόριθμος (1) για υπολογισμό του πίνακα του LCS εξηγείται απλά:

**Αν στο i,j είναι vi = wj τότε S(i,j) = S(i-1,j-1)+1**

Αλλιώς **S(i,j) = max( κελί της προηγούμενης σειράς, κελί προηγούμενης στήλης )**

Την οπισθοδρόμηση και την επιλογή των κελίων αφού φτιάξουμε τον πίνακα θα την κάνουμε ως εξής:

Ορίζουμε i=n,j=m και φτιάχνουμε ένα κενό string μήκους D(n,m).

Σε επανάληψη while για όσο i,j >0 :

Αν S1(i) = S2(j) τότε βάζουμε αυτό το στοιχείο στη τελευταία θέση του lcs και μειώνουμε κατά ένα τα i,j.

Aν δεν είναι ιδιά τότε κοιτάμε ποιο στοιχείο είναι μεγαλύτερο -προηγούμενη σειρά ή προηγούμενη στήλη και μειώνουμε το αντίστοιχο i ή j χωρίς να βάζουμε τιμή στο lcs.

Στο τέλος παίρνουμε το τελικό lcs το οποίο “χτίζεται” από το τέλος προς στην αρχή.

**Ο κώδικας:**

def lcs(Sm, Sn):

  m = len(Sm)

  n = len(Sn)

# Creation of the matrix

  DP = [[0 for x in range(n+1)] for x in range(m+1)]

  #rows

  for i in range(m+1):

    #columns

    for j in range(n+1):

      if i==0 or j==0:

        DP[i][j] = 0

      elif Sm[i-1] == Sn[j-1]:

        DP[i][j] = DP[i-1][j-1] +1

      else:

        DP[i][j] = max(DP[i-1][j], DP[i][j-1])

# Backtracking

  index = DP[m][n]

  lcs = [""] \* (index+1)

  lcs[index] = ""

  i=m

  j=n

  while i>0 and j>0:

    if Sm[i-1] == Sn[j-1]:

      lcs[index-1] = Sm[i-1]

      i -= 1

      j -= 1

      index -= 1

    elif DP[i-1][j] > DP[i][j-1]:

      i -= 1

    else:

      j -= 1

  print("S1: ", Sm)

  print("S2: ", Sn)

  print('Length of LCS: ', len(lcs))

  print("LCS: ", "".join(lcs))

Ερώτημα 6:

**Α)**

Θα δημιουργήσουμε τον πίνακα δυναμικού προγραμματισμού και για κάθε στοιχείο θα ακολουθήσουμε τον κανόνα:

Εάν vi == wj:

S(i,j) = max( S(i-1,j)-ρ, S(i,j-1)-ρ, S(i-1,j-1)+1 )

Εάν vi != wj:

S(i,j) = max(S(i-1,j)-ρ, S(i,j-1)-ρ, S(i-1,j-1)-ρ-σx)

Αν τα ρ,σ είναι θετικοί βλέπουμε ότι η μέθοδος δεν προτιμάει να υπάρχουν 2 συνεχόμενα mismatches καθώς δίνει μεγάλη ποινή.

Μαζί με τον πίνακα Σ θα φτιάξουμε έναν Χ ιδίων διαστάσεων τον οποίο θα αρχικοποιήσουμε με μηδενικα και θα αποθηκευουμε τις τιμες του x και έναν πινακα b στοn οποιο θα αποθηκευσουμε τους pointers για όταν κανουμε οπισθοδρομιση.

**Κώδικας:**

*# start*

m=len (S1)

n = len(S2)

*# creation of x with 0*

**For** i <- 0,m+1 **do**

**For** j <- 0,n+1 **do**

X(i,j) = 0

*# fill first row and column*

**For** i <-1,n **do**

S(i,0) = S(i-1,0) -ρ

**For** i <-1,m **do**

S(0,i) = S(0,i-1) -ρ

*# Create S matrix with all the values*

**For** i <-1, n **do**

**For** j <-1, m **do**

Flag = 0

x = X(i-1,j-1)

**If** vi == wj **then**

S(i,j) = max( S(i-1,j)-ρ, S(i,j-1)-ρ, S(i-1,j-1)+1 )

**If** S(i,j) = S(i-1,j-1)+1 **then**

B(i,j) = “\”

Flag = 1

**Else**

S(i,j) = max( S(i-1,j)-ρ, S(i,j-1)-ρ, S(i-1,j-1)-ρ-σx )

**If** S(i,j) == S(i-1,j-1)-ρ-σx **then**

X(i,j) = x+1

B(i,j) = “\”

Flag = 1

**If** flag == 0 **then**

**If** S(i,j) = S(i-1,j) **then**

B(i,j) = “|”

**Else**

B(i,j) = “<--”

*# print the strings Aligned*

i = len(S1)

j = len(S2)

**While** i > 0 or j > 0 **do**

V = [] *# S1*

W = [] *# S2*

**If** bi,j == “\” **then**

V.append( S1i )

W.append( S2j )

**Else** if bi,j == “|” **then**

V.append( S1i )

W.append( “-” )

**Else** # “<--”

V.append( “-” )

W.append( “S2j” )

i = i -1

j = j -1

V = V.reverse

W = V.reverse

*# finish*

Ερώτημα 8:

Εδώ θα χρησιμοποιήσουμε το πρόγραμμα του ερωτήματος 7 για να υπολογίσουμε το LCS των ακολουθιών SARS-CoV-2 και Bat-RaTTG13 τους όποιους κατεβάζουμε από εδώ και τους βάζουμε σε 2 αρχεία .txt:

<https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?report=fasta>

<https://www.ncbi.nlm.nih.gov/nuccore/MN996532.2?report=fasta>

Φορτώνουμε τις ακολουθίες τις μετατρέπουμε σε string και αφαιρούμε τους χαρακτήρες \n που υπάρχουν αναμεσά τους.

Το πρόγραμμα εκτελείται σε περίπου 10 λεπτά και δεσμεύει περίπου 12 GB RAM, οπότε η εκτέλεση δεν πραγματοποιήθηκε τοπικά αλλά σε cloud.

Η μέγιστη κοινή υπο-ακολουθία έχει μήκος 28751.

Για τον SarsCov-2 με μήκος 29903 στοιχεία αντιστοιχεί στο 96.1% του μήκους του και

Για τον Bat-RATTG13 με μήκος 29855 στοιχεία αντιστοιχεί στο 96.3% του μήκους τους.

Παρατηρούμε ότι οι ακολουθίες μοιάζουν σε μεγάλο βαθμό.

Εδώ φαίνεται η μέγιστη κοινή υποακολουθία**:**

ATTAAAGGTTTATACCTTCCAGGTAACAAACCAACAACTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACTTTAAAATCTGTGTGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAATAACTAATTACTGTCGTTGACAGGACACGAGTAACTCTCTATCTTCTGCAGGTGCTTACGGTTTCGTCCGTGTTGCAGCCGATCATCAGCACATCTAGGTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTTCAACGAGAAAACACACGTCCAACTCAGTTTGCCTGTTTACAGGTTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGGTTTTGCCTCAACTTGAACAGCCCTATGTGTTCATCAAACGTTCGATGCTCGAACTGCACCTCATGGCATGTTATGGTTGAGCTGGTAGCAGAACTAAGGCATTCAGTAGGTCGTAGTGGTGAGACACTGGTGTCCTTGTCCCTATGTGGGCGAAAACCAGTGGTTACCGCAAGGTTCTTCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCGATCTAAAGTCTTTGACTTAGGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAACTGGAACACTAAACATAGCAGTGGTGTACCCGTGACTCATGCGTGAGCTTAAGGAGGGCATACACTCGCTATGTCGATAACAACTTCTGTGGCCCTGATGGCTACCCTCTTGAGTGCATTAAAGACCTTCTAGCCGTGCTGGTAAAGCTTCATGCACTTTGTCCGAACAACTGGACTTTATTGACACTAAAGGGTGTATACTGCTGCGTGAACATGAGCATGAATTGCTTGGTACACGGAACGTTCTGAAAAGAGCTATGAATTGCAGACACCTTTTGAAATTAAATTGGCAAAGAAATTTGACACTTAATGGGGAATGTCCAAATTTTGTATTCCCTAAATTCCAAATCAAGACTATTCAACCAAGGGTTGAAAAGAAAAAGCTTGATGGCTTTATGGGTAGAATTCGATCTGTCTATCCAGTTGCTCACCAAATGAATGCAACCAAATGTGCCTTTCAACTCTCATGAAGTGTGATCATTGTGGTGAAACTTCATGGCAGACGGCGATTTTGTTAAAGCCACTTGGAATTTTGTGGCACTGAAATTTGACTAAAGAGGTGCCACTACTTGTGGTTACTACCCCAAAATGCTGTTGTAAAATTTATTGTCCAGCATGTCAAATCAGAAGTAGGACCTGAGCATAGTCTTGCGAATACATAATGAATCTGGCTTGAAACCATTCTTCGTAAGGGTGGTCGCACTATTGCTTTGGAGGCTGTGTGTTCTCTATGTTGGTTGCTACAAGTGTGCCTATTGGTTCCACGTGCTAGCGCAACATAGGTTGAACATACAGGTGTTGTTGGAGAAGGTTCGAAGGTCTTAATGAAACCTTCTTGAAATACTCAAAAAGAGAAAGTCAACATCAATATTGTTGGTGACTTTAAACTTAATGAAGAGATCGCCATTATTTTGGCATCTTTTTCTGCTTCACAAGTGCTTTTGTGGAAACTGTGAAAGGTTTGGATTATAAACATTCAAACAAATTGTTGAATCCTGGGTAATTTAAAGTTACAAAGGAAGCAAAAAAGGTGCTGGAAATTGGTGAACAAAATCAATATGAGTCCTCTTTATGCATTTGCTCAGAGGCTGCTCGTGTTGTCGATCAATTTTCTCCCGCACTCTTGAAACTGCTCAAAATTCGTGCGTGTTACAGAAGGCCGCTATAACAATACTAGATGGAATTTCACAGTATCACTGAGACTCATTGATGCTATGATGTTCACATCTGATTTGGTACTAACAATCTGTTGTAATGGCTACATTACAGGTGGTGTTGTTCAGTTGACTTCCAGTGGCTAACAAATCTTTGGCACTGTTTATGAAAACTCAAACCGTCTTGATTGGCTGAAGAGAATTAAGGAAGGTGTAGAGTTTCTTAGAGAGGTTGGGAAATTGTTAAATTTATCTCAACTGTGCTTGTGAAATTGTCGGTGGACAAATTGTCACCTGTGCAAAGGAAATTAAGAGAGTGTTCAGACATTCTTTAAGCTTGTAAAAAATTTTTGGCTTTTGTGCTGACTCATCATATTGGTGGAGCTAAACTTAAAGCCTTGAATTTAGGTGAAACATTTGTCACCACTCAAAGGGATTGTAAGAAAGTGTGTTAAACCAAGAAGAAACTGGCTACTCATGCCTCTAAAGCCCAAAAGAAATTATTTCTTAGAGGGAGAAACACTTCCACAGAAGTGTTAACAGAGGAAGTTGTCTTGAAAACTGGTGATTTACAACCATTAGAACAACCTACTAGTGAAGCTGTTGAAGCCCATTGGTTGGTACACCAGTTTGATTAACGGGCTATGTTGCTCGAAATAAAGACACAGAAAAGTACTGTGCCCTTGCACCTAATATGATGGTAACAAACAATACCTTCACACTAAAGGCGGTGCACCAACAAAGTTACTTTTGGTGATGACACTGTGATGAAGTGCAAGGTTAAAGAGTGTGAATATCACTTTTGAACTTGAGAAAGGATTGATAAGTACTTAAGAGAAGTGCTCTCCTATACAGTTGAACTCGGTACAGAAGTAAATGAGTTGCTGTGTTGTGCAGATGCTGTCATAAAAACTTTCAACCAGTATCTGAATTACTTAACCACTGGGCATTGATTTAGATGAGTGGGTATGGCTACATACTACTTATTTGATGAGTCTGGTGATTTAAATTGGCTTCACATATGTATGTTCTTTTACCTCCAGATGAGGATGAGAAGAAGGTGATTGTGAAGAAGAAGATTTGACCACAACTCAATATGAGTATGGTACTGAAGATGATTACCAAGGTAAACTTTGGAATTTGGTGCCACTTCTGTCCTTCAACCTGAAGAAGAGCAGAAGAAGATTGGTTAGATGATGATAGTCAACAAACTGTGTCAAAAGAGCAGTGAAAATCAGACAACTATATCAACAATTGTGAGGTTCAACCTCATTAGAGATGGAACTACACCAGTTGTTCAGACTGAAGTGAATAGTTTAGTGGTTATTTAAAACTTACTGACAATGTATACATTAAAAATGCAGACATTGTGGAAGAAGCTAAAAAGGTAAAACCAACATGTTGTAAGCAGCCAATGTTTACCTTAAACATGGAGGAGGTGTTGCGGAGCTTAAATAAGGCTACTAAAATGCCATGCAAGTTGAATCGATATTACATAGCACAATGGACCACTTAAAGTGGGTGGTGTGTGTTTTAGGGACACAATCTTGCTAAAACTGTCTTCATGTTGTGGCCAAATGTTAACAAGGTGAAGAATTCAACTTCTTAAGAGTGCTTATGAAAATTTTAATCAGCACGAGTTCTACTGCACCATATTATCAGCGGTATTTTTGGTGCTGACCCTTACATTCTTTAAGAGTTTGTGTAGAACTGTTCGCACAAATGTCTACTAGCTGTCTTTGATAAAAATCTCTATGACAAACTTGTTTCAAGCTTTTTGAAATGAAGAGTGAAAAGCAAGTTGAACAAAAGACGCTGAATTCCTAAAGAGGAAGTTAAGCATTATACTGAAAGTAAACTTCAGTTGACAGAGAAACAAGTGATAAGAAAATCAAAGCTTGTGTTGAAGAAGTTACAACAACTCTGGAAGAAACTAAGTTCTCACAGAAAATTGTTACTTATATTGACATTAATGGCAATCTTCATCCAGATTCTGCCACTCTTGTTATGACATTGACATCACTTTCTTAAAGAAAGATGCTCCTATATAGTGGGTGATGTTGTTCAAGAGGGTGTTTTAACTGCTGTGTTATACCTACTAAAAAGGCTGGTGGACTACTGAAATGCTACGAAAGCTTGAGAAAAGTGCCAACAGAAATATATAACCACTTACCCGGCAGGGTTTAAATGGTTAACTGTAGAGAGGCAAGACAGTGCTTAAAAAGTGTAAAAGTGCCTTTTACATTCTACCATCTATTATCTCTAATGAGAAGCAAGAAATTCTTGGAACTGTTTCTTGGAATTTGCGAGAAATGCTTGCACATGCAGAGAAACCGAAATTAATGCCTGTCTGTTGGAAACTAAAGCATAGTTTCAACTATACAGCGTAAATAAAGGGTATTAAAATACAAGAGGGTGTGGTTGATTATGGTGCAGATTTTACTTTTAACCAGTAAAACAACTGTAGCTCACTTATCAAACACTAACGATCTAAATGAACTCTTGTTACAATGCCACTTGGCTATGTAACACATGGTTAAATTTGGAAGAAGCTGCTCGGTATATGAGATCTCTCAAAGTGCCAGCTACAGTTTCTGTTTCTTCACCTGATGCTGTTACAGCGTATAATGGTTACTTACTTCTTCTTCTAAAACACCTGAAGAACATTTTATTGAAACCATCTCACTTGCTGGCCTTAAAGATTGGTCCTATTCTGGACATCTACACAACTAGGTATAGAATTTCTTAAGAGAGGTGATAAAAGTGTTATTACACTAGTAACCTAACATTCACCTAGATGGTGAAGTTATCACCTTTGACAATCTTAAGAACTTCTTCTTTGAAGAAGTAGGACTATTAAGGTGTTTACAACAGTAGACAACATTAACCTCACACGCAAGTTGTGGACATGTCAATGACATATGGACAACAGTTTGGCCAACTTATTTGGATGGAGCTGATGTTACTAAAATAAAACCTCATAATTCACATGAAGGTAAAACTTTTATGTTTTCCTAATGATGACACTTACGTGTGAGGCTTTTGAGTACTACCACACAACTGATCCTAGTTTTCTGGGTAGGTACATGTCAGCATTAAATCACACTAAAAAGTGGAAATACCCACAAGTAATGGTTTAACTTCTATTAAATGGGCAGATAACAACTGTTATCTTGCCACTGCATTTTAACACTAACAAATAGAGTTGAATTTAACCACCTGCTTACAAGATGCTATACAGGCAAGGCTGGTGAAGCTGCTAATTTTGTGCACTTATCTTAGCCTACTGTAAAAGACAGTAGGTGAGTTAGGTGATGTTAGAGAAACAATGATTATTGTTTCAACATGCCAATTTAGATTCTTGCAAAAGAGTCTTGAAGTGGTGTGTAAAACTTGTGGACAACAGCAGACACCTTAAGGGTGTAGAAGCTGTTATGTACATGGGCACACTTTCTTATGAACAATTAAGAAGGTGTTCAGATACCTTGTACTGTGGTAAACAAGCTACAAATATCTAGTACAACAGAGTCACCTTTTGTTATGATGTCGCACCACCTGCTCAGTATGAACTTAAGCATGGTACATTTATTGTGCTAGTGAGTACACTGGTAATTACCAGTGTGGTCACTATAAACATATAACTTCTAAAGAAACTTTGTATTGCATAGACGGTGCTTTACTTACAAAGTCCTCGATACAAAGGTCCTATTACGGATTTTTCTACAAAGAAAAAGTAAAACAACCATAAAACCAGTTATTATAAATTGGATGGTGTTGTTTGTACAGAAATTGACCTAAGTTGGACAATTATTATAAGAAAGACAATTCTTATTTCACAGAGCAACCAATTGATCTTGTCCAAACCAACCTATCCAAGCAAGCTTCGATAATTTAAGTTTGTATGTGATAAATCAAATTGCTGATGATTTAAACCAGTTAACTGGTTAAAGAAACCTGCTTCAAGAGAGTTAAAGTTACATTTTCCCTGACTTAAATGGTGATGTGGTGGCTATTGATTATAAACACTACACACCTCTTTTAAGAAAGGAGCTAAATTGTTACATAAACCTATTGTTGGCATGTTAACAATGCAACTAAAAAGCACGTATAAACCAAACCTGGTGTATACGTGTCTTTGGACACAAAACCAGTTGAAACATCAAATCTTTGATGTACTGAATCAGAGGACCGCAGGGAATGGATAATCTTGCCTGCGAAGATCTAAAACCAGTCTCTGAAGAAGTAGTGGAAAATCCTACCATACAGAAAGACTTCTTGAGTGTAATGTGAAAACTACCGAAGTTGTAGGAGACATTATACTTAAACCAGCAAATATGTTTAAAATTACAGAGAGGTTGGCACACAGATCTAATGGCTGCTTATGTAGACAATTCTAGTCTTACTATTAAGAAACCTAATGAATTATCTAGAGTATAGGTTTGAAAACCCTTGTACTCATGGTTTAGCTGCTGTTAATAGTGTCCTGGGATACTATAGCTAATTATGCTAAGCCTTTTCTTAAAAAGTTGTTAGACAACTACTAAATAGTACACGGTGTTTAAACCGTGTTTGTACTAATTATATGCCTTATTTCTTTACTTTATTGCTACAATTGTGTACTTTTACTAGAAGTACAAATTCTAGAATTAAAGCATCTATGCCGACTACTATAGCAAAGAATACTGTTAAGAGTGTGGTAAATTTTGCTAGAGGCTTCATTTAATTATGAAGTCACCTAATTTTTTAAATGATAATATTATAATTTGGTTTTTACTATTAAGTGTTTGTAGGTTCTTTAATCTACTCAACCGCTGCTTAGGTGTTTTAATGTCTAATTTAGGCATGCCTTCTTACTGTACTGGTACAGAGAAGGCTATTTGAACTCTACTAATGTCACTAGCAACTACTGTACTGGTTCTATACCTTGTAGTGTTTGTCTTAGTGGTTTAGATTCTTTAGAACCTATCCTTCTTTAGAAACTATACAAATTACCATTTCATCTTTTAAATGGGATTTAACTGCTTTTGGCTTAGTTGCAGAGTGGTTTTTGGCATATATTCTTTTACTAGGTTTTCTATGTACTTGGATTGGCTGCAATATGCAATTGTTTTTCAGCTATTTTGCAGTACATTTTATTAGTAATTCTTGGCTTATGTGGTTAATAATTAATCTTGTACAAATGGCCCCATTTCAGCTATGGTTAGATGTACATTTCTTTGCATCATTTTATTATGTATGGAAAAGTTATGTGCATGTTGTAGAGGTTGTATTCATCAACTTGTATGATGTGTTACAAACGTAATAGAGCAACAAGAGTGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCCTTTTATGTCTATGCTAATGGAGGTAAAGGCTTTTGCAAACTACAAATTGGAATTGTGTTAATTGTGAACATTCTGTGCTGGTAGTACATTTATTAGTGATGAAGTTGCGAGAGACTTGTCACTACAGTTTAAAAGACCAATAAATCCTACTGACCAGTCTTCTACATGTTGAAGTGTTACAGTGAAGAATGGTTCCATCCATCTTTACTTTGATAAAGCTGGTCAAAAGACTTATGAAAGACATTCTCTCTCTCATTTTGTTAACTTAGACAATGAGAGCTATAACACTAAAGGTTCATTGCCTATTAATGTTATAGTTTTTGATGGTAAATCAAAATGTGAAGAATCATCTGCAAAATCAGCGTCTGTTTATACAGTCAGCTTATGTGTCAACCTATACTGTTACTAGATCAGGCATTAGTGTCTGATGTTGGTGATAGTCGAAGTTGCAGTTAAAATGTTTGATGCTTACGTTAATACGTTTTCATCAACTTTTAACGTCCAATGGAAAAACTAAAACACTAGTTGCAACTGCAGAAGCTGAACTTGCAAAGAATGTGTCCTTAGACAATGTCTTATCACTTTATTTCAGCAGCTCGGCAAGGGTTTGTTGATTCAGATGTAGAAACTAAAGATGTTGTTGAATGTCTTAAATTGTCACATCAATCTGACATAGAAGTTACTGGGATAGTTGTAATAACTATATGCTCACCTATAACAAAGTTGAAAACATGACACCCGTGACCTTGGTGCTTGTATTGACTGTAGTGCGCGTCATATTAATGCGCAGGTAGCAAAAAGTCACAAATTGCTTGATATGGAACGTTAAAGATTTATGTCATGTCTGAACAACTACGAAAACAAATACGTAGTGCTGCAAAAAGAAAACTTCCTTTAAGTTGACATGTGCAACTACTAGACAAGTTGTTAATGTTGTAACAACAAAATAGCACTTAAGGGTGGTAAAATTGTTAATAATTGGTTGAAGCAGTTAATTAAAGTTACACTGTGTTCCTTTTTGTTGCTGCTATTTCTATTTAATAACACCTGTTCATGTCATGTCTAAACATACTGACTTTTCAAGTGAAATATAGGTACAAGGCTATGATGGTGGTGTCACTCGTGACATAGCATCTACAGATACTTGTTTTGCTAACAAACATGCTGATTTTGACACATGGTTTAGCAGCGTGGTGGTAGTTATACTAATGACAAAGCTTGCCATTGATTGCTGCAGTCATAACAAGAGAAGTGGGTTTTGTCGTGCCTGGTTTCCTGGACGATATTACGCACAACTAATGGTGACTTTTTGCATTTCTTACCTAGAGTTTTTAGTGCAGTTGGTAACATCTGTAACACCATCAAAACTTATAGAGTACACTGACTTTGCAACTCAGCTTGTGTTTTGGCTGCTGAATGTACAATTTTTAAAGATGCTTCTGGTAAGCCAGTACCATATTGTTATGATACAATGTACTAGAAGGTTCTGTTGCTATGAAAGTTACGCCTGAACACGTTATGTCTCATGGATGGCTCTATTATTCAATTTCCTAACACCTACCTTGAAGGTTCTGTTAGAGTGGTAACAACTTTTGATTCTGAGTATGTAGCACGGACTTGTGAAAGATCAGAAGCTGGTGTTTGTGTATCTACTAGTGGTAGATGGGTACTTAAAATGATTATTACAGATCTTTACCAGGAGTTTTTGTGGTGTAGATGCTGTAAATTTACTTACTAATATGTTACACCACTAATTCAACCTATGGTGCTTTGGACATATCAGCATCTATGTAGCGGGGTATTGTGCTATCGTAGTAACATGCCTTGCCTACTATTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGCATGTAGTTGCCTTTAATACTTTACTATTCCTTATGTCATTCACTGTACTCTGTTTAACACCAGTTTACTCATTTTACCTGGGTTTATTCTGTTATTTACTTGTACTTGACATTTTATCTTACTAATGATGTTTCTTTTTAGCACATATCAGTGGATGGTTATGTTCACACCTTTAGTACCTTTCTGGATAACAATTGTTTATCATTTGTATTTCCACAAAGCATTTTTTGGTTCTTTAGTAATACCTAAAAACGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGTTGCACTTTTTTTAAATAAGAAATGTATCTAAATTGCGTAGTGATGTCTTTACCTCTTACGCAATATAATAGATATTAGCCTTTATAATAAGTACAAGTATTTTAGTGGAGCATGGATACAACTAGCTACAGAGAAGCTGCTTGTTGTCATCTGCAAAGCTCTCAATGATTCAGTAATCAGGTTCTGATGTTCTTTACAACCACCACAAACCTCTATCACCTCGCTGTTTTGCAGAGTGGTTTTAGAAAAATGGCATTCCCATCGGTAAAGTTGAGGGTTGATGGTACAAGTACTTGTGGTACAACACACTTAAGGTCTTTGGCTTGATGAGTAGTTACTGTCCAAGACATGTGATCTGCACCTCTGAAGACATGCTTAACCTAATTATGAAGATTACTTTCGAAGTCTAATCATAATTTCTTGGTACAGGCTGGTAATGTCAACTAGGTTATTGGACATTCTATGCAAAATTGTGTCTTAAGCTTAAGGTGATACGCCAATCTAAGACCCTAAGTATAAGTTTGTTCGCATTCAACCGGACAGACTTTTTCAGTGTTAGCTTGTTACAATGGTTCACCATCTGGTGTTTACCATGTGCTATGAGCCAATTTACATTAAGGGTTCATTCCTTAATGGTTCTGTGGTAGTGTTGGTTTTAACATAGATTATGACTGTGTCTCTTTTTGTTACATGCACCATATGGAATTACCAACTGGAGTTCATGCTGGCACAGACTTAGAAGGTACTTTATGGACCTTTTGTTGACAGCAAACAGCACAAGCAGCTGGTACGGACACAACTATACAGTAATGTTTTAGCTTGGTTGTACGCTGCTGTTATAAATGGAGACAGGTGGTTTCTCAATCGATTTACCACAACTCTTAATGATTTAACCTGTGGCTATGAAGTACAATTATGAACCTCTAACACAGACCATGTTGACATACTAGGACCTCTTTCTGCTCAAACGGAATTGCCGTTTTAGATATGTGTGCTTCATAAAAGAATTACTGCAAAATGGTATGAATGGACGTACCATATTGGGTAGTGCTTTATTAGAAGATGAATTTACACCTTTTGATGTTGTTAGACAATGCTCAGGTGTTACTTTCCAAAGTGCAGTAAAAGAACAATCAAGGGACACACCATGGTTGTTACTACAATTTTGACTTCACTTTTAGTTTTAGTCCAGAGTACTCAATGGTCTTTGTTCTTTTTTTGTATGAAAATGCCTTTTTCCTTTTGCTATGGGTATTATTGCTATGTCTGCTTTTGCAATGATGTTTGTAAACATAAGCATGCATTCTCTGTTTGTTTTGTTACCTTCTCTTGCACTGTAGCTTATTTAATATGGTCTATATGCCTGCTAGTTGGGTGATGCGATTATGACATGGTTGGATATGGTTGATACTAGTTTGTCTGGTTTAAGCTAAAAGACTGTGTTATGTATGCATCAGCGTAGTGTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGGTGTGGACACTTATGAATGTCTTGACACTCGTTATAAAGTTTATTATGGTAATGCTTTAGATCAAGCATTTCCATGTGGGCTCTTATAATCTCTGTTACTTCTAACTACTCGGTGTAGTTACAACTGTCATGTTTTGGCCAGAGGTATTGTTTTTATGTGTGTTGAGTATTGCCCTATTTTCTTCATAACTGGTAATACACTTCAGTGTATAATGCTAGTTTATTGTTTCTTAGGCTATTTTGTACTTGTTACTTTGGCTCTTTTGTTTACTCAACCGTACTTTAGACTGACTCTTGGTGTTTATGATTATTGTTTCACACAGAGTTTAGTATATGAATTCACAGGGATCTCCCACCAAGAATAGCATAGATGCCTTAAACTCAACATAAATTGTTGGGTGTTGGGGAAACCTTGTATAAAGTAGCCACTGTACAGTCTAAAATGTCAGATGTAAAGTGCACTCAGTAGTCTTACTCTCAGTTTTGCAACAACTAGAGTAGAATCATCATCTAAATTGTGGGCTCAATGTGTCAGTTACACAATGACATCTCTTAGCTAAAGATACTACTGAAGCCTTTGAAAAAATGGTTTCACTACTTTCTGTTTGCTTTCCATGCAGGGTGCTGTAGACATAAACAAGCTTTGTGAAGAAATGCTGACAACAGGGCAACCTTACAAGCTATAGCCTCAGAGTTTAGTTCCTTCCATCATATGCAGCTTTTGCTACGCTCAAGAAGCTTATGAGCAGGCTGTTGCTAATGGTGATCTGAAGTTGTCTTAAAAAGTTGAAGAAGTCTTTGAATGTGGCTAAATCTGAATTTGACCGTGATGCAGCCATGCAACGAAGTTGGAAAAGATGGCTGATCAAGCTATGACCCAAATGTATAAACAGGCTAGATCTGAGACAAGAGGGCAAAAGTTACTAGTGCTATGCAGACAATGCTTTTCACTATGCTTAGAAAGTTGGATAATGATGCACTCAACAACATTATCAACAATGCAAGAGATGGTTGTGTTCCTTGAACATAATACCTCTTACAACAGCAGCCAAATAATGGTTGTCATACCAGACTATAACACATATAAAAATACGTGTGATGGTACAACATTTACTTATGCATCGCATTGTGGGAAATCCAACAGGTTGTGATGCAGATAGTAAAATTGTTCAACTTAGTGAAATTAGATGGACAATTCACCTAATTTAGCATGGCCTCTTATTGTAACAGCTTTAAGGGCCAATTCTGCTGTCAAATTACAGAATAATGAGCTTAGTCCTGTTGCACTACGACAGATGTCTTGTGCTGCCGGTACTACACAAACTGCTTGCACTGATGACAATGCGTTAGCTTACTACAACACAACAAAGGGAGGTAGGTTTGTACTTGCATGTTATCCGATTTACAGGATTTGAAATGGGCTAGATTCCCTAAGAGTGATGGAACTGGTACTATCTATACAGAACTGGAACCACCTTGTAGGTTTGTTACAGACACACCTAAAGGTCCTAAAGTGAAGTATTTATACTTTATTAAAGGATTAAACAACTAAATAGAGGTATGGTACTTGGTAGTTTAGCTGCACAGTACGTTACAAGCTGGTAATGCAACAGAAGTGCCTGCCAATTCAACTGTATATCTTTCTGTGCTTTTGCTGTAGATGCGCTAAAGCTACAAAGATATCTAGCTAGTGGGGGACAACCAATCACTAATTGTGTTAAGATGTTGTGTACACACACTGGTACTGGTCAGGCAATAACAGTTACACCGAAGCCAATATGGATCAAGAATCCTTTGGTGGTGCATCGTGTTGTCTGTACTGCGTTGCCACATAGATCATCCAAATCCTAAAGGATTTTGTGACTTAAAAGGTAAGTATGTACAAATACCTACACTTGTGCTAATGACCCTGTGGGTTTTACACTTAAAAACACAGTCTGTACCGTCTGCGGTATGTGGAAAGGTTATGGCTGTAGTTGTGATCAACTCCGCGAACCCATGCTTCAGTCAGCTGATGCACAATCGTTTTTAAACGGGTTTGCGGTGTAAGTGCAGCCCGTCTTACACCGTGCGGCACAGGCACTAGACTGATGTCGTTACAGGGCTTTTGACATCTACAATGATAAAGTAGCTGGTTTTGCTAAATTCTAAAAACTAATTGTTGTCGCTTCCAAGAAAAGACGAGATGACAATTTAATTGATTCTTACTTTGTAGTTAAGAGACACACTTTCTCTAACTACCAACATGAAGAAACAATTTATAATTTACTTAAGGATTGTCCAGCTGTTGCTAAACAGACTTCTTTAAGTTTAGAATAGACGGTGACATGGTACCACATATATCACGTCAACGTCTTACTAAATACACAATGGCAGACCTCGTCTATGCTTTAAGGCATTTTGAGAAGGAATTGTGAACATTAAAGAAATACTTGTCACATACAATTGTTGTGAGATGATTATTTCAATAAAAAGACTGGTATGATTTTGTAGAAAACCCAGATATATTACGCGTATACGCCAACTTAGGTGAACGTGTACGCCAAGCTTTGTTAAAAACAGTACAATTCTGTGATGCCATGCGAATGCTGGTATTGTGGTGTACTACATTAGATAATCAAGATCTCAATGGTAACTGGTATGATTTCGGTGATTCATACAAACCACCCAGGTAGTGGAGTTCCTTTGTAGATTCTTATTATTCATTGTTAATGCCTATATAACTTACAGGGCTTAACTGCAGAGTCACATGTTGACACTGACTTAACAAAGCCTACATTAAGTGGGATTTGTTAAAATATGACTTACGGAAGAGAGGTTAAAACTCTTTGACCGTTATTTTAAATATTGGGATCAGACATACCACCCAAATTGTGTTAACTGTTTGGATGACAGATGCATTCTGCATTGTGCAAACTTTAAGTTTTATTCTCTACAGTGTTCCCACCTACAAGTTTTGGACCCTAGTGAGAAAAATATTTGTTGATGGTGTTCCATTTGTAGTTTCAACTGGATACCACTTCAGGAGCTAGGTGTTGTACATAATCAGGATGTAAACTTACATAGCTCTAGACTTAGTTTTAAGGAATTACTTGTGTATGCTGCTGACCCTGCTATGCACGCTGCTTCTGGTAATTATTACTAGATAAACGCACTACGTGCTTTCAGTAGCTGCACTTACTAACAATGTTGCTTTTCAAACTGTCAAACCCGGTAATTTTAACAAAGACTTCTATGACTTTGCTGTGTCTAAGGGTTTCTTTAAGGAAGGAAGTTCTGTTGAATTAAAACACTTCTTCTTTGCTCAGGATGGTAATGCTGCATCAGCGATTATGACTACTATCGTTATAATCTACCAACAATGTGTGATATCAGACAACTCTATTTGTAGTTGAAGTTGTTGATAAGTACTTTGATTGTTACGATGGTGGCTGTATTAATGCTAACCAAGTCATCGTCAACAACCTAGACAAATCAGCTGGTTTTCCATTTAATAAATGGGGTAAGGCTAGACTTATTAGATTCAATGAGTTAGAGGATCAAGATGCACTTTTCGCATATACAAAACGTAATGTCATCCCTACTATAACTCAAATGAATCTTAAGTATGCCATTAGTGCAAAGAATAGAGCTCGCACCGTAGCTGGTGTCTCTATCTGTAGTACTATGACCAATAGACAGTTTCATCAAAAATTATTGAAATCAATAGCCGCCACTAGAGGAGCTACTGTAGTAATGGAACAAGCAAATTTATGGTGGTTGGCAAACATGTTAAAAACTGTTTAAGTGATGTAGAAAACCCTCACTTATGGGTTGGGATTACCTAAATGTGATAGAGCCATGCCTAACATGCTTAGAATTATGGCCTCACTTGTTCTTGCTCGCAAACATACAACGTGTGTAGCTTGTCACACCGTTTCTATAGATTAGCTAATGAGTGTGCTCAAGTATTGAGTGAAATGGTCATGTGTGGCGGTTCACTATATGTTAAACCAGGTGGAACCTCATCAGGAGATGCCACAACTGCTTATGCTAATAGTGTTTTAACATTTGTCAAGCTGTACGGCCAATGTTAATGCACTTTTATCTACTGATGGTAACAAAATTGCCGATAAGAGTCCGCAATTTACAACACAGACTTTATGAGTGTCTCTATAGAAATAGAGATGTTGACACAGACTTTGTGAATGAGTTTTACGCATATTTGCGTAAACATTTCTCAATGATGATACTTCTGAGATGCTGTTGTGTGTTTCAATAGCACTTATGCATCTCAAGGTCTAGTGGCTAGCATAAAGAACTTTAATCAGTTCTTTATATCAAAACAAGTTTTTATGTCTGAAGCAAAATGTTGGACTGAGACTGACCTTACTAAAGGACCTCATGAATTTTGCTCTCAACATACAATGCTAGTTAAACAGGGTGATGATTATGTGTACCTCCTTACCCGATCCATCAGAATCTAGGGGCGGCTGTTTTGTAGATGATATCGTAAAAACAGATGGTACACTATGATTGAACGGTTGTGTCTTTAGCTATAGATGCTTACCCACTTACTAAACATCCTAATCAGGAGTATGCTGATGTCTTTCATTGTACTTACAATACATAAGAAAGTACATGATGAGTTAACAGGACAATGTTAGACATGTATTCTGTTATGCTTACTAATGATAACACTTCAAGGTATTGGGAACCTGAGTTTTATGAGGCTATGTACACACCGCATACAGTCTTACAGGCTGTTGGGGCTTGTGTTCTTTGCAATTCACAGACTTCATTAAGATGTGGTGCTTGCATACGTAGACCATTCTTATGTGTAAATGCTGTTACGACCAGTCATATCACATCACATAAATTAGTCTTGTCTGTTAATCCGTATGTTTGCAATGCTCCAGGTTGTGATGTCACAGATGTGACTCAACTTTACTTAGGAGGTATGAGCTATTATTGTAAATCACATAAACCACCCATTAGTTTTCCTTGTGTGCTAATGGACAAGTTTTTGGTTTATATAAAATACATGTGTTGGTAGCGATAATGTTACTGACTTTAAGCAATTGCAACATGTGATGGACAAATGCTGGTGATTACATTTTAGCTAACACCTGTACTGAAAGACTCAAGCTTTTGCAGCAGAAACGCTCAAAGCTACTGAGGAGACATTTAAACTGTCTTATGGTATTGCACTGTACGTGAAGTGCTGTCTGACAGAGAATTCATCTTTCATGGGAAGTTGGTAAACCTAGACCACCACTTAACCGAAATTATGTCTTTACTGGTTATCGTGTAACTAAAAACAGTAAAGTACAAATAGGAGAGTACACCTTTGAAAAAGGTGACTATGGTGATGCTGTTGTTACCGAGGTACAACAACTTACAAATAAATGTTGGTGATATTTTGTGCTGACATCACATACAGTAATGCCATTAAGTGCACCTACACTAGTGCCACAAGAGCACTATGTTAGAATTACTGGCTTATACCCAACACTCAATATTCAGATGAGTTTTCTAGCAATGTTGCAAATTATCAAAAGGTTGGTATGCAAAAGTATTCTACACTCCAGGACCACCTGGTACTGGTAAGAGTCATTTTGCTATTGGCTAGCTCTCTACTACCCTTCTGCTCGCATAGTGTAACAGCTTGCTCTCATGCCGCTGTTGATGCACTATGTGAGAAGGCATTAAAATATTTGCCTATAGATAAATGTAGTAGAATTATACCTGCACGTGCTCGTGTAGAGTGTTTTGATAAATTCAAAGTGAATTCAACATTAGAACAGTAGTCTTTTGTACTGTAAATGCATTGCCTGAGACGACGCAGATATAGTTGTCTTTGATGAAATTTCAATGGCACAAATTATGATTTGAGTGTTGTCAATGCAGATTACGTGCTAAGCACTATGTGTACATTGGCGACCCTGCTCAATTACCGCACCACGCACATTGCTAACTAAGGGCACACTAGAACCAGAATATTTCAATTCAGTGTGTAGACTATGAAAACTATAGGTCCAGACATGTTCCTGGAACTTGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAGCTAAAGCACATAAAGACAAATCAGCTCAATGCTTTAAAATGTTTTATAAGGGTGTTATACCATGATGTTTCATCTGCAATTAACAGGCCACAAATAGGCGTGGTAAGAGAATTCTTACACGTAACCTCTTGGAGAAAAGCTGTCTTATTTCACCTTATAATTCACAGAATGCTGTAGCCTCAAAGATTTTGGGACTACCAACTCAAACTGTTGATTCATCACAGGGTCAGAATATGACTATGTCATATTCACTCAAACCACTGAACAGCTCACTCTTGTAATGTAAACAGATTTAATGTTGCTATTACAGAGCAAAAGTAGGCATACTTTGCATAATGTCTGATAGAGACCTTTATGACAAGTTGCAATTTACAAGTCTTGAATTCCACGTAGGAATGTGGCAACTTTACAAGCTGAAAATGTAACAGGACTTTTAAGATTGTAGTAAGGTAATCACTGGGTTACACCACACAGGCACCTACACACCTCAGTGTTGACACTAAATTCAAAACTGAAGGTTTATGTGTTGACATACCTGGATACCTAAGGACATGACCTATAGAAGACTCATCTCTATGATGGGTTTAAAATGAATTACAAGTTAATGGTTACCCTAAATGTTTATCACCCGGAGAAGCTAAGACATGTACGTGCATGGATTGGCTTGATGTCGAGGGGTGTCATGCTACTAGAGAAGCTTTGGTACAATTTACCTTTACAGCTAGGTTTTCTACAGGTGTTAACCTAGTTGCTGTACCTACAGGTATGTTGATACACCTAATAATACAGATTTTTCCAGAGTTAGTGCTAAACCACCCCTGGAGATCATTTAAACATCATACCACTTATGTAAAAGGACTTCCTTGGAATGTAGTGCGTATAAAGATTGTACAAATGTTAAGTGAACACTTAAAAATCTCTCTGACAGAGTCGTTTTGTCTTATGGGCACATGGCTTTGATTACATCTATGAAGTATTTTGTGAAAATAGGACCTGAGCGCACTGTTGTCTATGTGATAACGTGCCACATGCTTTTCCACTGCTTCAGACACTTATGCCTGTTGGCATCATTCTATTGGATTTGATTAGTCTATAATCCGTTTATGATTGATGTTCAACAATGGGGTTTTACAGGTAACCTACAAAGCAACCATGATCTGTATTGTCAGTCATGGTAATGCACATGTAGCTAGTTGTGATGCAATCATGACTAGTGTCTAGCTGTCCACGAGTGCTTTGTTAAGCGTGTTGACTGGACTATTGAATACCTATAATTGGTGATGAACTGAAGATTAATGCGGCTTGTAGAAAGGTTCAACACATGGTTGTTAAAGCTGCATTATTAGCAGACAAATTCCAGTTCTTCAGACATTGGTAACCCTAAAGCTATTAAGTGTGTACCTCAAGCTGATGTAGAATGGAAGTTCTATGAGCACAGCCTTGTAGTGACAAAGCTTATAAAATAGAAGAATTATTCTATTCTTATGCCACACATTCTGACAAATTCACAGATGGTGTATGCCTATTTTGGAATTGCAATGTCGATAGATATCCTGCTAATTCCATTGTTTGTAGATTTGACACTAGAGTGCTATCTAACCTTAACTTGCCTGGTTGTGATGGTGGCAGTTTGTATGTAAATAAACATGCATTCCACACACCAGCTTTTGATAAAAGTGCTTTTGTTAATTTAAAACAATTACCATTTTTCTATTACTCTGACAGTCCATGTGAGTCTCATGGAAAACAAGTAGTGTCAGATATAGATATGTACCACTAAAGTCTGCTACGTGATAACACGTTGCAATTTAGGTGGTGCTGTCTGTAGACATCATGCTAATGAGTACAGATTGTATCTGATGCTTAAACATGATGATCTCAGCTGGCTTTAGCTTGTGGGTTTACAAACAATTTGATACTTAAACCTCTGGAAACTTTTACAAGACTTCAGAGTTTAGAAAATGTGGCTTTTAATGTTGTAAATAAGGACACTTTGATGGACAACAGGGTGAAGTACCAGTTTCATCATTAATAACACTGTTTACACAAAAGTTGATGGTGTTGATGTAGAATTTTGAAAATAAAACAACATTACCTGTTAATGTAGCATTTGAGCTTGGGCTAAGCGCAAATTAAACCAGTACCAGAGGTGAAAATACTCAATAATTTGGGTGTGGACATTGCTGCTAATACTGTGATTGGGACTACAAAAGAGATGCTCCAGCACATATATCTACTATTGGTGTTTGTTCTATGACTGACATAGCCAAGAAACCAACTGAAAATTTGTGCACCACTCACTGTCTTTTTGATGGTAGAGTATGGTCAAGTAGACTTTTTAGAAATGCCCGTAATGGTGTTCTTATTACAGAAGGTAGTGTTAAAGGTTTACAACCATCTGTAGGTCCAAACAAGCAGTCTTAATGGAGTCACATTATTGGAGAAGCCTAAAAACACAGTTCAATTATTAAAGAAAGTTATGGTGTTGTCAACAATACCTGAAACTTACTTTACCAAGTAGAAATTTAAAGAATTAAACCCAGGAGTCAAATGGAAATTGATTTCTTAGAATTAGCTATGGATGAATTCATTGAACGGTATAAATAGAAGGTATGCCTTCGAACATATCGTTATGGAGATTTTAGTCATAGCAGTTAGGTGGTTTACATCTACTGATTGGACTAGCTAAACGTTTAAGGAATCACCTTTGAATTAGAAGATTTTATTCCTATGGACAGTACAGTTAAAAATTTTCATAACAGATGCGCAAACAGGTTCATCTAAGTGTGTGTGTTCTGTTATTGATTTATTACTTGATGATTTTGTTGAAATAATAAAATCCAGATTTATCTGTAGTTTCTAAGGTGTCAAAGTGACATTGACTATACAGAAATTTCATTTATGCTTTGGTGTAAAGATGGCATGTGAAACATTTTACCCAAAATTACAATCTAGTCAAGCGTGGCAACCGGGTGTTGCTATGCCTAATCTTACAAAATGCAAAGAATGTTTAGAAAAGTGTGACCTTCAAAATTATGGTGATAGTGCAACATTACCTAAAGGCATAATGATGAATGTCGCAAAATATACTCAACTGTGTCATATTTAAAACATTACTTAGCTGTACCCTATAATATGAGGTTATACATTTTGGTGCTGGTTCTGATAAAGGAGTTGCACCGGTACAGCTGTTTTAAGACAGTGGTTGCCTACGGGTACCTCTTGTCGATTCAGATCTTAATGACTTTGTCTCTGATGCAGATTCAACTTTGATTGGTGATTGTGCAACTGTACATACAGCTAATAAATGGGATCTCATTATTAGTGATATGTACGACCTAAGACTAAAAATGTTACAAAAGAAAATGATTAAAGAGGGTTTTTCACTTACATTTGTGGTTTATACAACAAAAGCTAGCCTTGGAGGTTCTGGCTATAAAGATAACAGACATTCTTGGAATGCGATCTTTATAAGCTATGGGACATTCGCATGGTGGACGCTTTGTTACTAATGTAATGCTCTCATCTGAAGCATTTTTAATTGGTGTAATTACTTGGCAAACCGCGAACAATAGATGGTTATGTCATGCATGCAAATTACATATTTTGGAGGAATACAAACCAATTCAGTTGTCTTCCTATTCTTTATTTGACATGAGTAAATTCCCCTTAAATTAAGGGGTACTGCGTTATGTCTTTAAAGAAGGTCAAATCAATGATATGATTTTATCTCTTCTTAGTAAAGGTAGACTTATAATTAGAGAAAACAAAGAGTTGTTATTTCTAGTGATGTTCTTGTTAACAACTAAACGAACATGTTTGTTTTTCTTGTTTTATTGCCACTAGTTCTAGTCAGTGTGTTAATCTACAACAGAACTCATTACCCCTGCATACACAACTCACACGTGGTGTTATTACCCTGACAAAGTTTTCAGATCTCAGTTTTACATTAACTCAGGATTGTTTTACCTTTCTTTCCAATGTACTGGTTCCATGCTATACATGTTTGGGACCAATGGTATAAAGGTTTGATAACCCGTCTCCATTAAGATGGGTTATTTTGCTTCCACTGAGAAGTCTAAATAATAAGAGGTGGATTTTTGGTACTACTTAGATTCGAAGACCCAGTCCTACTTATTGTTAATAACGCTACTAATGTTGTTATTAAAGTCTGTGAATTTCAATTTTGTAATGATCCATTTTTGGGTGTTTATTACCACAAAAACAACAAAAGTTGGATGGAAAGTGAGTTCAGAGTTTATCTAGTGCGAATAATTGCACTTTTGATATGTCTCTCAGCCTTTTCTTATGGACCTTGAAGGAAAACAGGGTAATTTCAAAAATCTTAGGGAATTGTGTTTAAGAATATTGATGGTTATTTAAAATATATTCTAAACACGCCTATTAATTTAGTGCGTGATCTCCCCGGTTTTTCGCTTTAGAACCATTGGTAGATTGCCAATAGGTATTAACATCACTAGGTTTCAAACTTTACTTGCTTTACATAGAAGTATTTGACTCCTGGTGATTCTTCTTCAGGTTGGACAGCTGGTGCTGCAGCTTATTATGTGGGTTATCTTCAACCAGGACTTTTCTATAAAATATAATGAAATGGAACCATTACAGATGCTGTAGACTGTGCACTTGACCCTCTTCAGAAACAAAGTGTACGTTAAATCCTTCACTGTGAAAAAGGAATTATCAAACTCTAACTTTAGAGTCCAACCAACAGATCTATTGTTAGATTCCAATATTACAAACTTTGCCTTTTGGTGAAGTTTTTAACGCCACCAATTGCATCGTTTATGCTTGGAACAGAAGAGAATAGCAACTGTGTTGCTGATTATCTGTCCTATATAATTCCATCATTTTCACTTTAATGTTATGGAGTGTCTCCTACTAAATTAAATGATCTCTGCTTTACTAATGTTATGCAGATCATTTGTATTAAGGTGATGAAGTCAGACAAATGCCCAGGCAAACTGGAAAGATTGCTGATAAATTATAAATACCAGATGATTTTACGGTGGTTATAGCTTGGAATTCTAACATTTGATCAAGGGGGGTAATTTAATACTTTAGTTTTTAGAAGCTAATCTAAACCTTTGAGAGGATATTCAACTGAAATTTCAGCGGAGCAACCTTGTAATGGTAAGGTTAATTGTACTCCTTTATATATGGTTTACCCACTATGGTGTTGGTACCAACCTAAGGTAGTAGTACTTTCTTTTGAACTTCTAATGCACCAGCAACTGTTTGTGGACCTAAAAGTCTACTAATTGGTTAAAAAAAATGTGTCAATTTCAACTTAATGGTTTAACGGCACAGGTGTCTACGAGTCTAAAAAAAGTTTCTCCTTTCCAACAATTTGGAGAGACATTGCGACACTACTGATGCGTCCGTGATCCACAGACACTTGAGATTCTTGACATTACACCATGTTCTTTTGGTGGTGTCAGTGTTATAACACCGGAACAAATCTCTAACCAGGTTGCTGTTCTTTATCAGGATGTTAACTGCACAGAAGTCCCTGTTGCTATCATGCAGACAACTTACTCCACTTGGCGTGTTTACTACAGGTTCTAATGTTTTTCAAACACGTGCAGGTGTTTAATAGGGGCTGAACATGTCAAAACTCTATGAGTGTGACATACCATTGGTGCAGGATATGCGCAGTTATCAGACTCAACTAATTCACGTAGTGTGCAGTCAATCATATTGCCTACACTATGTCACTTGGTGCAGAAAATTCAGTTGCTTATCTAATAACTCTATTGCCATACCACAAATTTTACTATTAGTGTACCACGAAATTCTACCGTGTCTATGACAAGACATCGTAGATGTACAATGTAATTTGTGGTGATTCAACTGATGCAGCAACTTTTGTTGCAATATGGAGTTTTTGACACAATTAAACGTGCTTTAACTGGAATAGCTGTTGAACAGACAAAAAACCAAGAAGTTTTTGCCAAGTAAACAAATTTAAAACACCACCAATTAAAGATTTTGGTGGTTTAATTTTTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGGTCATTTATTGAGATTACTTTTCAAAAAGTGACACTTGCGATGCTGGCTTCATCAAACAATATGGTGATTGCCTTGGTGATATTGCTGCTAGGATCATTTGTGCCAAAAGTTAAGGCCTTACTGTTTGCCACCTTTGCTCACAGATGAAATGATGCTCAATACACTTCTGCACTTTAGCGGTACAATCACTTCTGGTTGGACTTTGGTGCAGGTGCTGCTTACAAATACCATTTGCATGCAAATGGCTTATAGGTTTAATGGTATTGGAGTTACACAGAATGTTCTCTATGAGAACCAAAAATTGATTGCCAACCATTTAATAGTGCTATTGGCAAAATTCAAGACTCACTTTCTTCACAGCAAGTGCACTTGGAAAACTTCAAGATGTGTCAACCAAAATGCACAAGCTTTAAACACGCTTGTTAAACAACTTAGCTCCAATTTTGGGCATTTCAGGTTTAAATGATATCCTTTCACGTCTGACAAAGTTGAGGCTGAAGTGCAATTGAAGGTTGATCACAGGCAGACTTCAAAGTTGCAGACATATGTGACTCAACAATTAATTAGAGCTGCAGAAATCAGAGCTTCTGCAATCTTGCTGCTACTAAAATGTCAGAGTGTGTACTGGACAATCAAAAAGAGTTGATTTTTGTGGAAAGGCTATCATCTTATGTCTTCCCTCAGTCAGCACCTCATGGTGTAGTCTTCTTGCATGTGACTATGTCCCTGCACAAGAAAAGAACTTCACAACTGCTCCTGCCATTTGTCATGATGGAAAAGCACACTTTCCCGTGAAGGTGTTTTGTTTCAAATGGCACACACTGGTTTGTACACAAAGGAATTTTTATGAACCACAAATATTACACAGACAACACATTTGTTCTGGTACTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTTATGATCCTTTGCAACCGAATTGATCATTCAAGGAGGAGTTGATAAATATTTAAAATCATACATCACCGATGTGATTTAGGTGACATTCTGGCATTAATGCTTCAGTTGTAAATTCAAAAGAAATTGACCGCCTCAATGAGGTTGCCAAAATTAAATGAATCTCTCATCGATCTCCAAGAACTTGGAAAGTATGACAGTATATAAAATGGCCATGGTACATTTGGCTAGGTTTTATAGCTGGCTTGATTGCCATATAATGGTCAATTATGCTTTGCTGTATGACCAGTTGCTGAGTTGTCTCAAGGGCTGTTGTTCTTGGGATCTGCTGCAAATTTGATGAAGACGACTCTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACATAAACGAACTTATGGATTTGTTTATGAGAATTTCACATTGGAACTGTAACTTTGAACAAGGTGAAATAAGGATGCTACTCCTTCAGATTTGTTCGCGCTACTGCAACGATACCGATACAAGCCTCACTCCCTTTCGGATGGCTTATTGTTGGCGTTGCATTCTTGCTGTTTTTCAAGCGCTTCCAAATCATAACCCTAAAAAGAGATGGCAACTAGCACTCTCAAGGGTTTCACTTTTTTGCAACTTGCTGTGTGTTTGTAACAGTTTACTCACACTTTTGCTCGTTGCTGCTGGCTTGAAGCCCCTTCTCTACTTAGCTTTAGTCTACTTCTTGCAGAGTATAAACTTTGTAAGAATAATAATGAGGCTTTGGCTTTGCTGGAAATGCCGTTCCAAAAACCCATTACTTATGATGCAACTATTCTTTGTGGCATACTAATTGTTAGACTATTGTATACCTTACAATAGTGTAACTTCTTCAATTGTCATTACTTCAGGTGATGGCACAACAAGTCCTATTTCTGAACATGACTACCAGATTGGTGGTTATACTGAAAAATGGGATCTGGAGTAAAAGACTGTGTTGTATTACACAGTTACTTCACTTCAGATATTACCAGCTGTACTCAACTCAATTGAGACAGACACTGGTGTTGAACATGTTACCTTCTTCATCTACAATAAAATTGTTGATGAGCCTGAAGAACATGTCCAAATTCACACAATCGACGGTTCATCCGGAGTTGTTAATCCAGAATGGAACCAATTTATGATGAACCGACGACGACTACTAGCGTGCCTTTGTAAGCACAAGCTGATGAGTACGAACTTATGTACTCATTCGTTTCGGAAGAGACAGGTACGTTAATAGTTAATAGCGTACTTCTTTTTCTTGCTTTCGTGGTATTCTTGCTAGTACACTAGCCATCCTTACTGCGCTTCGATTGTGTGCGTACTGCTGCAATATTGTTAACGTGAGTCTTGTAAAACCTTCTTTTTACGTTTACTCTCGTGTTAAAAATCTGAATTCTTCTAGAGTTCCTGATCTTCTGGTCTAAACGAACTAAATATTATATTAGTTTTTCTGTTTGGAACTTTAATTTTAGCCATGGCAGATAACGGTACTATTACCGTTGAAGAGCTAAAAAGCTCCTTGAACATGGAACTAGTAATAGGTTCCTATTCTTACATGGATTTGTCTTCTACAATTTGCCTATGCCAACAGGAATAGGTTTTTGTATATAATTAAGTTAATTTTCCTCTGGCTGTTTGGCCAGTAACTTTAGCTGTTTGTGCTTGCTGCTGTTTACAGAATAAATTGGATCACGGGGAATGCTATCGCAATGGCTTGTCTTGTAGGCTTGATGTGGCTAGCTACTTCATTGCTTCTTTCAGCTTTTGCCGTACGCGTTCCATGTGGTCATTCAATCCAGAAACTAACATTTTCTCAACGTGCCACTCCATGGCACTATTTGACCAGACCGCTTCTAGAAGTGAACTGTAATCGGAGCTGTATCCTTCGTGGACATCTTCGTATTGCTGGACACCATCTAGGACGCTGTGACATCAAGGACCTGCCAAAGAAATCACTGTTGCTACATCACGAACGCTTTCTTATTACAAATTGGGAGCTTCCAGCGTGTAGCAGGTGATCAGGTTTTGCTGCATACAGTCGCTACAGGATTGGAACTAAAATTAAACACAGACCATTCCAGTAGCAGTGACAATATTGCTTTGCTTGTACAGTAAGTGACAACAGATGTTTCATCTCGTTGACTTTCAGGTTACTATAGCAGAGATATTACTAATTATTATGAGGACTTTAAAGTTTCCATTTGGAACTTGATTACATCATAAACCTATAATTAAAAATTTATCTAAGTCACTAACTGAGAATAAATATTCTCAATTAGATGAAGAGCAACCAATGGAGATTGATTAACGAACATGAAAATTATTCTTTTCTTGGACTGTAACACTGCTACTTGTGAGCTTTATCACTACCAAGAGTGTGTTAGAGGTACAACAGTACTTTAAAAGAACCTTGCTCTTCTGGAACTAGAGGCAATTCACCATTCATCCTCTAGCTGATAAAAATTTGCACTGACTTGCTTTAGCACTCAATTTGCTTTTGCTTGTCCTGACGGCGTAAACACGTCTATCAGTTACGTGCCAGATCAGTTTCACCAAACTGTTCATCAGACAAGAGGAAGTTCAAGAACTTTACTCCCAATTTTTCTTATTTGCGCAATAGTGTTTATAACACTTTGCTTCACACTCAAAAGAAAACAGAATGATGAACTTTCATTAATTGACTTCTATTTGTGCTTTTTAGCCTTTCTGCTATTCCTTGTTTTAATTATGCTTATTATCTTTTGGTTCTCACTTGAACTGCAAGATCATAATGAAACTTGTCACGCCTAAACGAACATGAAATTCTTGTTTTCTTAGGAATCTCACAACGTACTGCATTTCACCAAGAATGTAGTTTACAGTCATGTCTCAACACAACCATATGTAGTTGATGACCCGTGTCCTATTCACTTCTATTCTAAATGGTAATTAGAGTAGGAGCTAGAAAATCAGCACCTTTAATTGAATTGTGCGTGGATGAGGTGGTTCTAAATCACCCATTCAGTACATCGATATCGGAATTATACAGTTTCCTGTTACCTTTTACAATTAATTGCCAGGACCTAAATTGGGTAGTCTTGTAGTGCGTTGTTCGTTCTATGAAGACTTTTTAGAGTATCATGACGTTCGTGTTGTTTTAGATTTCATCTAAACGAACAAACTAAAATGTCTGATAATGGACCCCAAAACACGAAATGCACCCCGCATTACGTTTGGTGGACCCTCAGATTCAACTGGCAGTAACCAGAATGGAGAACGCAGTGGGCCGACAAAACAACGTCGGCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGGAAGACCTTAAATTCCCTCGAGGACAAGGCGTTCCATTAAACCAATAGCAGTCCAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGTAAAATGAAAGATCTCAGTCCAAGATGGTATTTTACTACCTAGGAACTGGGCCAGAAGCTGGACTTCCCTATGGTGCTAACAAAGACGGCATCATATGGGTTGCAACTGAGGGAGCCTTGAATACACCAAAAGATCACATTGGCACCCGCAATCCTGCTAACAATGCTGCAATCGTGCTACAACTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAGGGGCAGTCAAGCTCTTCTCGCCTCATCACGTAGTCGCAACAGTTCAAGAAATCAACTCCAGGCAGCAGTAGGGGAACTTCCCTGCTAGATGGCTGGCAATGGCGTGATGCTGCTCTTGCTTTGCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTCTGGTAAAGGCCAACAACAACAGGCCAAACTGTCACTAAGAAATCTGCTGCGAGGCTTCTAAGAACCTCGGCAAAAACGTACTGCCACAAACATACAATGTAACACAAGCTTTGGCAGACGTGGTCCAGAACAAACCCAAGGAAATTTGGGGACAGAACTAATCAGCAAGGAACTGATTACAAACATTGGCCGCAAATTGCACAATTCCCCCAGCGCTTCAGCTTCTTCGGAATGTCGCGCATTGGCATGGAAGTCACACCTTCGGGAACGTGGTTGACCTAACAGGTGCCATAAATTGGATGACAAAGATCCAAATTTCAAAGATCAAGTCATTTTGCTGAATAAGCAATTGACGCATACAAAACATTCCCACCAACAGAGCCTAAAAAGGACAAAAAGAAAAGGCTGATGAAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAACTGTGACTCTTCTTCCTGCTGCAGATTTGGATGATTCTCCAAACAATTGCAACAATCCATGAGCAGTGCTGATCAACTCAGGCCTAAACTCATGCAGACCACACAAGGCAGATGGGCTATATAAACGTTTTCGCTTTTCCGTTTACGATATATAGTCTACTCTTGTGCAGAATGAATTCTCGTAACTACATAGCACAAGTAGATGTAGTTAACTTAATCTCACATAGCAATCTTTAATCAGTGTGTAACATTAGGGAGGACTTGAAAGAGCCACCACATTTTCACCGAGGCCACGCGGAGTACGATCGAGGTACAGTGAAAATGCTAGGGAGAGCTGCCTATATGGAAGAGCCCTAATGTGTAAAATTAATTTTAGTAGTGCTATCCCATGTGATTTTAATAGCTTCTTAGGAGAATGACAAAAA

Ερώτημα 9:

1. Για το πρόβλημα της καθολικής ταύτισης θα χρησιμοποιήσουμε τον αλγόριθμο Needleman-Wunsch.

Έχουμε τις 2 συμβολοσειρές: V = GGTTCGTGGA, W = GGTTCGTGGA

Ο πίνακας Δυναμικού προγραμματισμού θα είναι:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **w\v** |  | **G** | **G** | **T** | **T** | **C** | **G** | **T** | **G** | **G** | **A** |
|  | 0 | -1 | -2 | -3 | -4 | -5 | -6 | -7 | -8 | -9 | -10 |
| **G** | -1 | 1 | 0 | -1 | -2 | -3 | -4 | -5 | -6 | -7 | -8 |
| **A** | -2 | 0 | 0 | -1 | -2 | -3 | -4 | -5 | -6 | -7 | -6 |
| **T** | -3 | -1 | -1 | 1 | 0 | -1 | -2 | -3 | -4 | -5 | -6 |
| **C** | -4 | -2 | -2 | 0 | 0 | 1 | 0 | -1 | -2 | -3 | -4 |
| **G** | -5 | -3 | -1 | -1 | -1 | 0 | 2 | 1 | 0 | -1 | -2 |
| **T** | -6 | -4 | -2 | 0 | 0 | -1 | 1 | 3 | 2 | 1 | 0 |
| **G** | -7 | -5 | -3 | -1 | -1 | -1 | 0 | 2 | 4 | 3 | 2 |
| **A** | -8 | -6 | -4 | -2 | -2 | -2 | -1 | 1 | 3 | 3 | 4 |
| **A** | -9 | -7 | -5 | -3 | -3 | -3 | -2 | 0 | 2 | 2 | 4 |
| **T** | -10 | -8 | -6 | -4 | -2 | -3 | -3 | -1 | 1 | 1 | 3 |
| **T** | -11 | -9 | -7 | -5 | -3 | -3 | -4 | -2 | 0 | 0 | 2 |

Score = 2  
Στοίχιση v: G G T T C G Τ G G A - -

w: G A T - C G T G A A T T

**ΙΙ)** Για το προβλημα της τοπικής αναζήτησης θα εφαρμόσουμε τον αλγόριθμο Smith-waterman. Αυτός είναι διαφορετικός στα σημεία όπου όλες οι πιθανές τιμές ενός κελιού είναι αρνητικές, στα οποία θα βάζουμε 0.

Στον πίνακα τις τιμές οι οποίες δεν προκύπτουν από τοπική ταύτιση αλλά από μετατόπιση θα τις βάζουμε κανονικά αλλά δεν θα βάζουμε βελάκια προκειμένου στο σχήμα να μείνουν μόνο οι τοπικές ακολουθίες καθώς θα εφαρμόσουμε τοπική ταύτιση χωρίς κενά.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **w\v** |  | **G** | **G** | **T** | **T** | **C** | **G** | **T** | **G** | **G** | **A** |
|  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **G** | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| **A** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 2 |
| **T** | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| **C** | 0 |  | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 |
| **G** | 0 | 1 | 1 | 0 | 0 | 1 | 3 | 2 | 1 | 1 | 0 |
| **T** | 0 | 0 | 0 | 1 | 1 | 0 | 2 | 4 | 3 | 2 | 1 |
| **G** | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 3 | 5 | 4 | 3 |
| **A** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 4 | 4 | 5 |
| **A** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 3 | 5 |
| T | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 2 | 2 | 4 |
| T | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 1 | 1 | 1 | 3 |

Ο πίνακας έχει μέγιστη τιμή 5 σε τρία κελιά και 5 τοπικές στοιχήσεις

Score = 5

V: GGT**TCGTG**GA

W: GA**TCGTG**AATT