Ad-Soyad : Email : No : İmza :

Final (04.06.2010, Süre: 60 dk

0114820 - Biyoenformatiğe Giriş

	S1	S2	S3	S4	S5	S6	Toplam		

S1. DHK amino asit dizisini sentezlemesi mümkün olan DNA dizileri hangileridir? Bulunuz.(Ekteki tabloyu kullanınız). (10)

		Seco	nd letter		1
	U	С	Α	G	
U	UUU } Phe UUA } Leu	UCU UCC UCA UCG	UAU Tyr UAC Stop UAG Stop	UGU Cys UGC Stop UGG Trp	UCAG
С	CUU CUC CUA CUG	CCU CCC CCA CCG	CAU His CAA GIn CAG	CGU CGC CGA CGG	UCAG
A	AUU AUC AUA lle AUG Met	ACU ACC ACA ACG	AAU Asn AAC AAA AAA Lys	AGU Ser AGC AGA Arg	UCAG
G	GUU GUC GUA GUG	GCU GCC GCA GCG	GAU Asp GAC GAA GAG	GGU GGC GGA GGG	UCAG

A=Ala=Alanine
C=Cys=Cysteine
D=Asp=Aspartic acid
E=Glu=Glutamic acid
F=Phe=Phenylalanine
G=Gly=Glycine
H=His=Histidine
I=Ile=Isoleucine
K=Lys=Lysine
L=Leu=Leucine
M=Met=Methionine
N=Asn=Asparagine
P=Pro=Proline
Q=Gln=Glutamine
R=Arg=Arginine
S=Ser=Serine
T=Thr=Threonine
V=Val=Valine
W=Trp=Tryytosine

(15)

\_\_\_\_\_\_

- S2. **SEVML** ve **EVCL** dizileri verildiğine göre;
  - a. İki diziye ait normal dot plotu çiziniz.
  - b. (a) da çizdiğiniz dotplotu kullanarak en uygun dizilişi nasıl elde ettiğinizi göstererek bulunuz. (10)

	S	E	V	M	L
E					
V					
C					
L					

c. bu iki diziye ait "local alignment score" tablosunu dinamik programlama algoritmasını kullanarak oluşturun (skorları belirlemek için BLOSUM62 matrisini kullanın, gap ceza puanı -3). ( $_{Hattrlatma:}$   $s_{i,j}$ = MAXIMUM[ $s_{i-1,j-1}$ + $s(a_i,b_j)$ ,  $s_{i,j-1}$ +w,  $s_{i-1,j}$ 

(05)

d. (c) de oluşturduğunuz matrisi kullanarak en uygun dizilişi bulunuz.

Α	4	l																		
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
С	0	-3	-3	-3	9															
Q	-1	1	0	0	3	5		_												
E	-1	0	0	2	4	2	5		_											
G	0	-2	0	7	ကု	-2	-2	6		_										
Н	-2	0	1	-1	ფ	0	0	-2	8											
I	-1	ფ	-3	-3	-1	-3	ფ	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	7	ო	1	1	-2	-1	ကု	-2	5								
M	-1	7	-2	ო	-1	0	-2	ო	-2	1	2	7	5							
F	-2	ფ	-3	ფ	-2	-3	ფ	-3	-1	0	0	ფ	0	6						
Р	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
s	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Т	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5		i	
w	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Υ	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
٧	0	ფ	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4
	Α	R	N	D	C	ø	Е	G	H	-	L	K	M	F	P	s	T	W	Y	٧

	S	E	V	M	L
E					
V					
C					
L					

BLOSUM62 amino asit substitution matrisi

\_\_\_\_\_\_

S3.	Veri Tabanı Sistemi (Database System) nedir? Kaç tür Veri Tabanı Sistemi vardır? Açıklayınız.	(10)
	rbirleri ile ilişkili dört farklı türe ait diziler ATCC, ATGC, TTCG, and TCGG olarak verilmiştir. Bu türlerin aralarınd olarak Hamming uzaklığı kullanan basit bir kümeleme algoritması kullanarak bu türlerin ilişkisini veren phylogeneti	laki farkların
S5. Aş	ağıda verilen Perl "one-liner" ne yapar, açıklayınız.	(5)
	\$ perl -npe 'last if $\d{4}$ \$/;' embl.data	
	φ peri inperiuse in γ (α ( ) γ φ γ, emericana	
S6.	Bu soruyu çözmek için <a href="http://www.yildiz.edu.tr/~naydin/na_Intro2Bio.htm">http://www.yildiz.edu.tr/~naydin/na_Intro2Bio.htm</a> sayfasına gidip ilgili linkten gerekli do indirmeniz gerekmektedir.	syayı (20)
	monneniz gerenitencom.	(20)

## **CEVAPLAR**

C1. Solution:

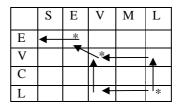
D=Asp=Aspartic acid = {GAU, GAC} H=His=Histidine = {CAU, CAC} K=Lys=Lysine = {AAA, AAG}

Possible RNA sequences	Possible DNA sequences
GAU CAU AAA	CTA GTA TTT
GAU CAU AAG	CTA GTA TTC
GAU CAC AAA	CTA GTG TTT
GAU CAC AAG	CTA GTG TTC
GAC CAU AAA	CTG GTA TTT
GAC CAU AAG	CTG GTA TTC
GAC CAC AAA	CTG GTG TTT
GAC CAC AAG	CTG GTG TTC

- C2. A database system is a computer program (or group of programs) that provides a mechanism to define and manipulate one or more databases.
  - Personal database systems: Designed to run on PCs
    - Access, Paradox, FileMaker, dBase
  - Enterprise database systems: Designed to support efficient storage and retrieval of vast amount of data
    - Interbase, Ingres, SQL Server, Informix, DB2, Oracle
  - Open source database systems: Free!!! (Linux!!!)
    - PostgreSQL, MySQL

C3.

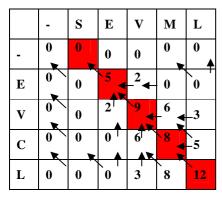
(b)



Dizilim 1:

n 1:						Dizilim 2:					
$\mathbf{S}$	$\mathbf{E}$	$\mathbf{V}$	M	_	L	S	$\mathbf{E}$	$\mathbf{V}$	_	M	L
_	E	$\mathbf{V}$	_	C	L	_	E	$\mathbf{V}$	C	_	L

(c)



E V M L
E V C L
$$5 + 4 + -1 + 4 = 12$$

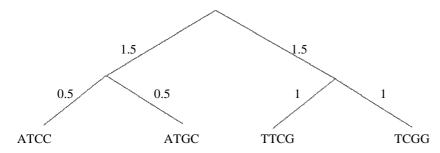
		ATCC	ATGC	TTCG	TCGG
•	ATCC	0	1	2	4
•	ATGC	1	0	3	3
•	TTCG	2	3	0	2
•	TCGG	4	3	2	0

	{ATCC,ATGC}	TTCG	TCGG
{ATCC,ATGC}	0	2.5	3.5
TTCG		0	2
TCGG			0

• İlk küme: { ATCC,ATGC}

Bir sonraki küme (cluster): {TTCG, TCGG}

Buna göre phylogenetic ağaç:



C5.

This is a Perl one-liner that prints only those lines from the embl.data disk-file that do not end in four digits.

## C6. Çözümlerden biri aşağıdaki gibidir:

```
open (infile, "list.txt");
open (outfile, ">list_out.txt");
$cnt = 0;
while ($line = <infile>){
        if ( $line =~ m/Adaptive/ || $line =~ m/adaptive/) {
            $cnt++;
            $nlcon =~s///g;
            $nlcon .= $line;
        }
}
```

print outfile "Total number of occurrences of the word "adaptive" =  $c. \n\n$ "; print outfile  $n.\n$ "; print outfile  $n.\n$ ";