Bioinformatika 1

Završni ispit - rješenja

14. lipnja 2021.

1. (5 bodova)

Zadana je Burrows-Wheelerova transformacija niza *S*: BWT(*S*) = GT$AAG. Rekonstruirajte originalan niz *S* iz zadane transformacije korištenjem LF-mapiranja. Postupak je potrebno skicirati. Pretpostavite da je znak $ abecedno manji od ostalih znakova abecede nad kojima je niz *S* izgrađen.

Rješenje:

Stupac L = BWT(*S*); stupac F su leksikografski sortirani znakovi iz stupca L.

Niz *S* se rekonstruira od zadnjeg znaka prema prvome.

S = AGTAG$

F	**L**
$	G
A	T
A	$
G	A
G	A
T	G

G$

F	**L**
$	G
A	T
A	$
G	A
G	A
T	G

AG$

F	**L**
$	G
A	T
A	$
G	A
G	A
T	G

TAG$

F	**L**
$	G
A	T
A	$
G	A
G	A
T	G

GTAG$

F	**L**
$	G
A	T
A	$
C	A
G	A
T	G

AGTAG$

2. (6 bodova)

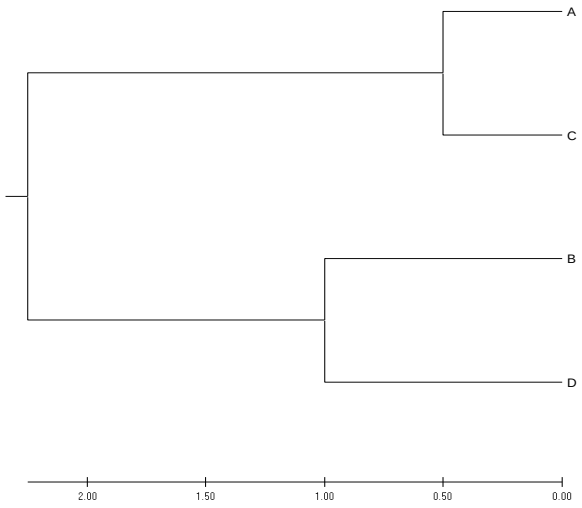
Izgradite filogenetsko stablo korištenjem metode UPGMA za zadanu ulaznu matricu udaljenosti. Potrebno je u svakom koraku izgradnje skicirati stablo i označiti izračunate udaljenosti između taksona. Matrica udaljenosti:

	A	**B**	**C**	**D**
A	0	4	1	3
B	4	0	6	2
C	1	6	0	5
D	3	2	5	0

Rješenje:

Ne vrijedi nejednakost trokuta za sve kombinacije taksona.

No, filogenetski alat MEGA generira sljedeće stablo (zanemaruju se preduvjeti):



3. (5 bodova)

Izračunajte entropije niza *S*: $H_0(S)$ i $H_1(S)$ za zadani niz *S* = CAAAC.

Rješenje:

$n = 5$, $n_c = 2$, $n_G = 3$

$H_0(S) = -(\frac{n_C}{n} log \frac{n_C}{n} + \frac{n_A}{n} log \frac{n_A}{n}) = -(\frac{2}{5} log \frac{2}{5} + \frac{3}{5} log \frac{3}{5}) = 0.97$

con_1 = C, S^{con1} = A

con_2 = A, S^{con2} = AAC

$H_0(S^{con1}) = 0$

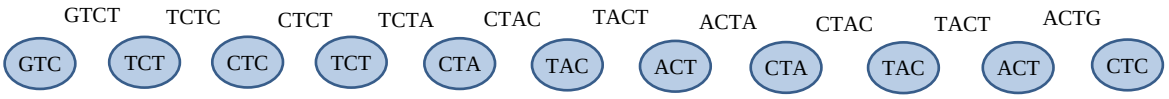
$H_0(S^{con2}) = -(\frac{1}{3} log \frac{1}{3} + \frac{2}{3} log \frac{2}{3}) = 0.918$

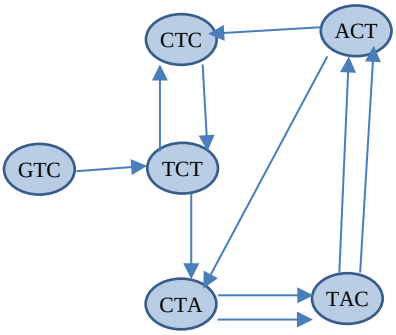
$H_1(S) = \frac{1}{n} \sum_{con \in \Sigma^1} |S^{con}| H_0(S^{con}) = 0.2 \cdot (|S^{con1}| \cdot H_0(S^{con1}) + |S^{con2}| \cdot H_0(S^{con2})) = 0.55$

4. (5 bodova)

Za niz s=GTCTCTACTACTC napraviti očitanja koristeći k-torke duljine **4** (k-torke predstavljaju niz uzastopnih nukleotida, a počinju sa svakim nukleotidom u nizu osim zadnjih k-1, npr. prva je GTCT). Na osnovi očitanja nacrtati pojednostavljeni de Bruijnov graf i pronaći sve Eulerove staze u njemu i na osnovu njih ispisati moguće izlazne nizove.

Rješenje:



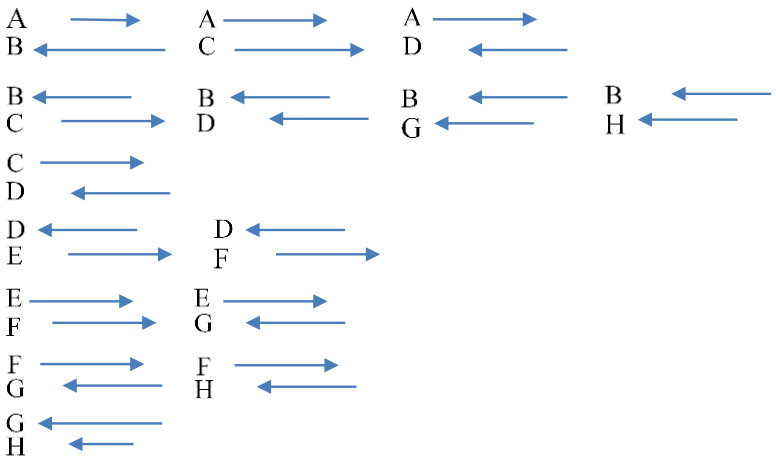


Nizovi:

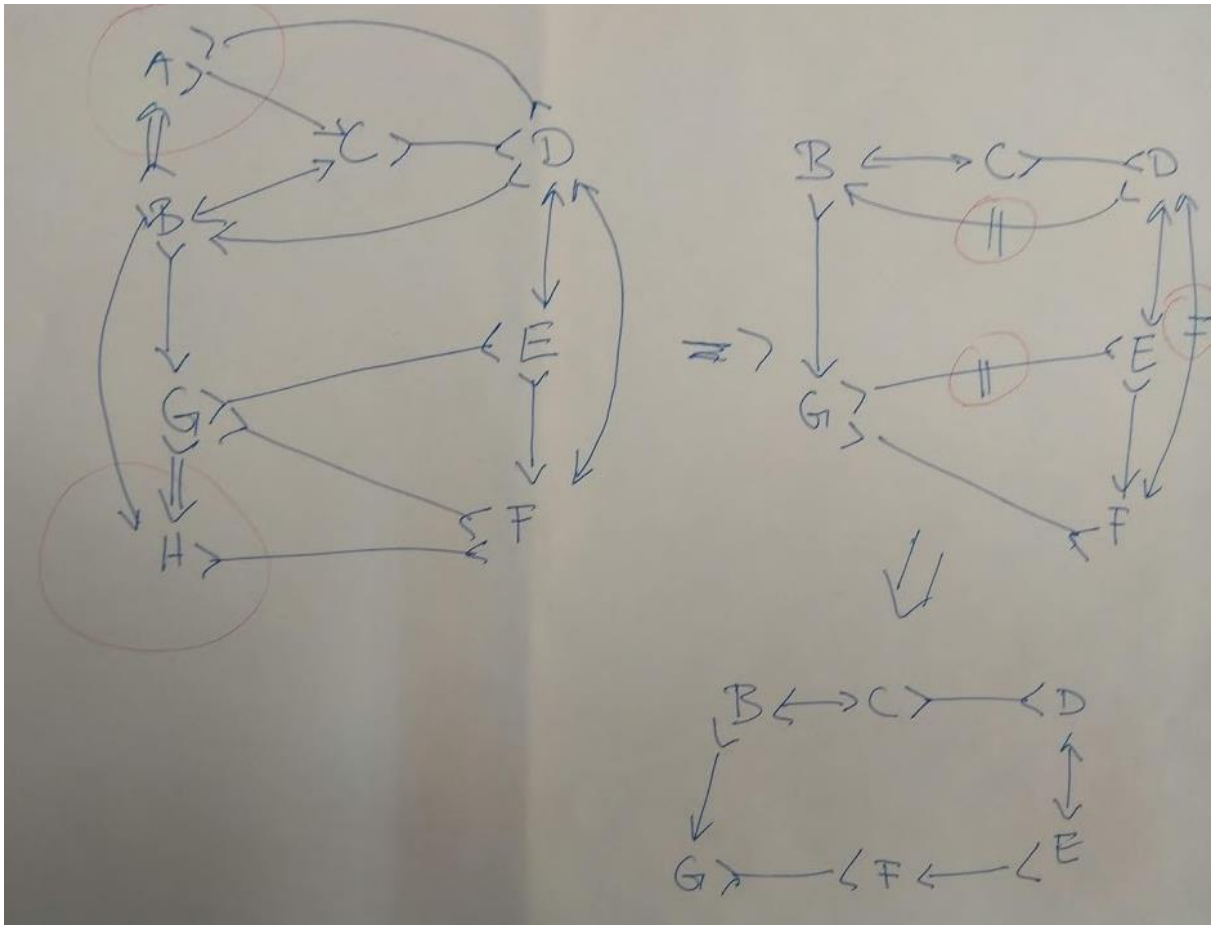
GTCTCTACTACTC
GTCTACTACTCTC

5. (5 bodova)

Za zadana preklapanja nacrtajte zajednički graf preklapanja te isti pojednostavite koristeći OLC pristup. Označite dobivene blokove.



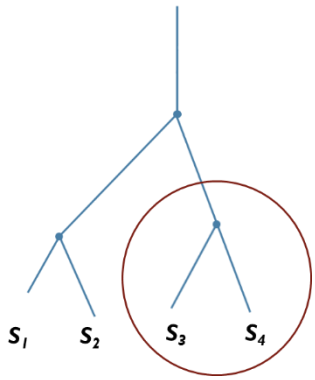
Rješenje:



6. (3 boda)

Skicirajte jedno ukorijenjeno filogenetsko stablo za četiri slijeda S_1, S_2, S_3, S_4. Prikažite na skiciranom stablu jednu monofiletsku skupinu te objasnite njezino značenje

Odgovor:



Monofiletsku skupinu čine organizmi (ili sljedovi) koji imaju istog pretka.

7. (2 boda)

Definirajte Eulerovu stazu i Hamiltonov put.

Odgovor:

Eulerova staza obilazi sve bridove grafa točno jednom, Hamiltonov put obilazi sve čvorove grafa točno jednom.

8. (2 boda)

Ako imamo N poravnatih nizova, na koji način određujemo konsenzus tih poravnanja?

Odgovor:

Težinskim/odnosno većinskim glasanjem nukleotida na svakoj pojedinoj poziciji u nizovima.

9. (2 boda)

Što je to pokrivenost ili dubina sekvenciranja (*engl. coverage*) u sastavljanju genoma i kako ga računamo?

Odgovor:

Prekrivanje se definira kao C = N L / G, pri čemu je N broj očitanja, L prosječna duljina svakoga očitanja, a G duljina segmenta.