| | |
|---|---|
| JMBAG | |
| Ime i prezime | |

# Bioinformatika

**Pismeni ispit**

**20. veljače 2020.**

**Rješenja 1.-5. zadatka napisati na vlastitim papirima, a odgovore na 6.-10. pitanje napisati na testu. Sve predati u košuljici.**

***1. (10 bodova)***

a) Za zadani niz *S* = ACTGGACT$ potrebno je nacrtati sufiksno stablo. Na sufiksnom stablu označite sufiksne veze. Pretpostavite da je znak $ leksikografski manji od ostalih znakova niza *S*.

b) Za zadani niz *S* potrebno je odrediti polje pokazivača na LMS-podnizove prema algoritmu SA-IS.

***2. (10 bodova)***

Odredite Burrows-Wheelerovu transformaciju niza *S* = ACTGGACT$. Pretpostavite da je znak $ leksikografski manji od ostalih znakova niza *S*.

Skicirajte traženje niza *P* = ACT u nizu *S* korištenjem LF-mapiranja.

***3. (10 bodova)***

Zadana su 4 slijeda: $S_1$ = ACT, $S_2$ = GTT, $S_3$ = GTA, $S_4$ = ACA. Nacrtajte sva moguća neukorijenjena stabla za ova 4 slijeda i odredite metodom najmanjeg broja evolucijskih promjena (eng. *maximum parsimony*) koje od mogućih stabala najbolje objašnjava filogenetske odnose između zadanih sljedova.

***4. (10 bodova)***

Prikažite lokalno poravnanje dva niza P = GATCTAG i S= CATGCTTC korištenjem Smith-Waterman algoritma (umetanje -1, brisanje -1, nepodudaranje -1, podudaranje 2,). Prikažite matricu poravnanja zajedno s početnim uvjetima i samo poravnanje. U matrici prikažite put poravnanja (pretraga unatrag).

***5. (10 bodova)***

Za niz s=GTCACCCAGACA napraviti očitanja koristeći k-torke (k-torke predstavljaju niz uzastopnih nukleotida, a počinju sa svakim nukleotidom u nizu osim zadnjih k-1, npr. prva je GTC). K =3. Na osnovi očitanja nacrtati pojednostavljeni de Bruijn graf i pronaći sve Eulerove staze u njemu i na osnovu njih ispisati moguće izlazne nizove.

***6. (2 boda)***
Objasnite što su BLOSUM matrice i čemu služe.

Odgovor:

---

***7. (2 boda)***
Napišite matematički izraz kojim se određuje nulta entropija niza *S* duljine *n*.

Odgovor:

---

***8. (2 boda)***
Navedite tri postupka za pojednostavljenje grafa u fazi razmještanja OLC pristupa za sastavljanje genoma?
Odgovor:

---

***9. (2 boda)***
Što je to Hamiltonov put, a što je to Eulerova staza?
Odgovor:

---

***10. (2 boda)***
Navedite najmanje složenosti (vremenske i prostorne) u O-notaciji koje možete postići algoritmima dinamičkoga poravnanja u slučaju kada morate znati optimalni put poravnanja za dvije sekvence duljine n (Napomena: nije poznata sličnost, odnosno udaljenost poravnanja pa nije moguće koristiti ograničeno dinamičko programiranje).

Odgovor:

---

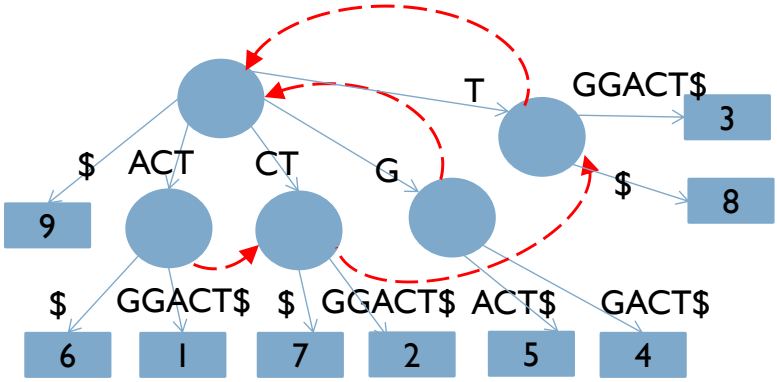
***1. (10 bodova)***

a) Za zadani niz *S* = ACTGGACT$ potrebno je nacrtati sufiksno stablo. Na sufiksnom stablu označite sufiksne veze. Pretpostavite da je znak $ leksikografski manji od ostalih znakova niza *S*.

b) Za zadani niz *S* potrebno je odrediti polje pokazivača na LMS-podnizove prema algoritmu SA-IS.

Rješenje:

a)




b)

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| ***S*** | A | C | T | G | G | A | C | T | $ |
| ***t*** | S | S | L | L | L | S | S | L | S |

| | | |
|---|---|---|
| $P_1$ | 6 | 9 |

***2. (10 bodova)***

Odredite Burrows-Wheelerovu transformaciju niza *S* = ACTGGACT$. Pretpostavite da je znak $ leksikografski manji od ostalih znakova niza *S*.

Skicirajte traženje niza *P* = ACT u nizu *S* korištenjem LF-mapiranja.

*S* = ACTGGACT$
BWT(*S*) = **TG$AAGTCC**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| A | C | T | G | G | A | C | T | $ |
| C | T | G | G | A | C | T | $ | A |
| T | G | G | A | C | T | $ | A | C |
| G | G | A | C | T | $ | A | C | T |
| G | A | C | T | $ | A | C | T | G |
| A | C | T | $ | A | C | T | G | G |
| C | T | $ | A | C | T | G | G | A |
| T | $ | A | C | T | G | G | A | C |
| $ | A | C | T | G | G | A | C | T |

→

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $ | A | C | T | G | G | A | C | **T** |
| A | C | T | $ | A | C | T | G | **G** |
| A | C | T | G | G | A | C | T | **$** |
| C | T | $ | A | C | T | G | G | **A** |
| C | T | G | G | A | C | T | $ | **A** |
| G | A | C | T | $ | A | C | T | **G** |
| G | G | A | C | T | $ | A | C | **T** |
| T | $ | A | C | T | G | G | A | **C** |
| T | G | G | A | C | T | $ | A | **C** |

Traženje podniza *P* = ACT

ACT

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $ | A | C | T | G | G | A | C | T |
| A | C | T | $ | A | C | T | G | G |
| A | C | T | G | G | A | C | T | $ |
| C | T | $ | A | C | T | G | G | A |
| C | T | G | G | A | C | T | $ | A |
| G | A | C | T | $ | A | C | T | G |
| G | G | A | C | T | $ | A | C | T |
| T | $ | A | C | T | G | G | A | C |
| T | G | G | A | C | T | $ | A | C |

A**C**T

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $ | A | C | T | G | G | A | C | T |
| A | C | T | $ | A | C | T | G | G |
| A | C | T | G | G | A | C | T | $ |
| C | T | $ | A | C | T | G | G | A |
| C | T | G | G | A | C | T | $ | A |
| G | A | C | T | $ | A | C | T | G |
| G | G | A | C | T | $ | A | C | T |
| T | $ | A | C | T | G | G | A | C |
| T | G | G | A | C | T | $ | A | C |

**A**CT

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $ | A | C | T | G | G | A | C | T |
| A | C | T | $ | A | C | T | G | G |
| A | C | T | G | G | A | C | T | $ |
| C | T | $ | A | C | T | G | G | A |
| C | T | G | G | A | C | T | $ | A |
| G | A | C | T | $ | A | C | T | G |
| G | G | A | C | T | $ | A | C | T |
| T | $ | A | C | T | G | G | A | C |
| T | G | G | A | C | T | $ | A | C |

***3. (10 bodova)***

***4. (10 bodova)***

Prikažite lokalno poravnanje dva niza P = GATCTAG i S= CATGCTTC korištenjem Smith-Waterman algoritma (umetanje -1, brisanje -1, nepodudaranje -1, podudaranje 2,). Prikažite matricu poravnanja zajedno s početnim uvjetima i samo poravnanje. U matrici prikažite put poravnanja (pretraga unatrag).

**Odgovor:**
Poravnanje
**AT-CT**
**ATGCT**

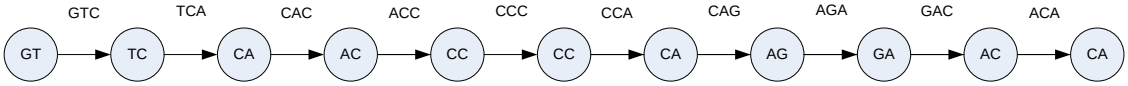
Matrica

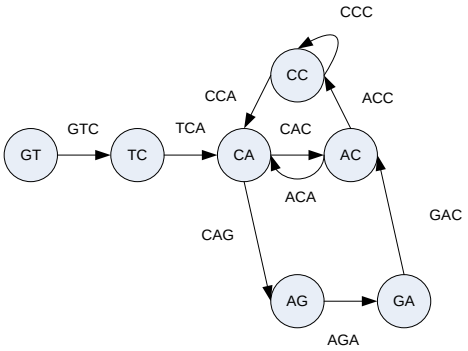
| | | C | A | T | G | C | T | T | C |
|---|---|---|---|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 |
| A | 0 | 0 | **2** | 1 | 1 | 1 | 0 | 0 | 0 |
| T | 0 | 0 | 1 | **4** | **3** | 2 | 3 | 2 | 1 |
| C | 0 | 2 | 1 | 3 | 3 | **5** | 4 | 3 | 4 |
| T | 0 | 1 | 1 | 3 | 2 | 4 | **7** | 6 | 5 |
| A | 0 | 0 | 3 | 2 | 2 | 3 | 6 | 6 | 5 |
| G | 0 | 0 | 2 | 2 | 4 | 3 | 5 | 5 | 5 |

***5. (10 bodova)***

Za niz s=GTCACCCAGACA napraviti očitanja koristeći k-torke (k-torke predstavljaju niz uzastopnih nukleotida, a počinju sa svakim nukleotidom u nizu osim zadnjih k-1, npr. prva je GTC). K =3. Na osnovi očitanja nacrtati pojednostavljeni de Bruijn graf i pronaći sve Eulerove staze u njemu i na osnovu njih ispisati moguće izlazne nizove.

**Rješenje:** Očitanja {GTC, TCA, CAC, ACC, CCC, CCA, CAG, AGA, GAC, ACA}




GTCACCCAGACA
GTCAGACACCCA
GTCACAGACCCA
GTCAGACCCACA

***8. (2 boda)***

Navedite tri postupka za pojednostavljenje grafa u fazi razmještanja OLC pristupa za sastavljanje genoma?

**Rješenje:**

- Izbacivanje sadržanih očitanja
- Izbacivanje tranzitivnih preklapanja
- Ujedinjavanje

***9. (2 boda)***

Što je to Hamiltonov put, a što je to Eulerova staza?

**Odgovor**: Hamiltonov put je put grafu koji obilazi sve vrhove u grafu točno jedanput.
Eulerova staza je staza u grafu koja obilazi sve bridove u grafu točno jedanput.

***10. (2 boda)***

Navedite najmanje složenosti (vremenske i prostorne) u O-notaciji koje možete postići algoritmima dinamičkoga poravnanja u slučaju kada morate znati optimalni put poravnanja za dvije sekvence duljine n (Napomena: nije poznata sličnost, odnosno udaljenost poravnanja pa nije moguće koristiti ograničeno dinamičko programiranje).

**Odgovor:** Vremenska složenost ($O(n^2)$) , prostorna složenost ($O(n)$) – Hirchbergov algoritam.