

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

**Poboljšanje djelomično
sastavljenog genoma dugim
očitanjima**

Lana Tuković, Ema Vlanić

Voditelj: *Krešimir Križanović*

Zagreb, svibanj 2023.

SADRŽAJ

1. Uvod	1
2. Opis algoritma	2
3. Zaključak	3

1. Uvod

Postupak sastavljanja složenih genoma može dati fragmentiran rezultat zbog velikog broja ponavljajućih sekvenci koje otežavaju proces poravnanja. Naš zadatak bio je pokušati međusobno povezati dobivene fragmente - contige u cijeli genom. Algoritam koji smo pri tome koristili zasniva se na konstruiranju grafa preklapanja te pronalaženja optimalnih staza među preklapanjima. Pronađena optimalna staza će nam služiti da povežemo dva contig-a.

2. Opis algoritma

Skupovi očitavanja i već sastavljenih contig-a pripremljeni su kao testni podaci. Njihova preklapanja dobivena pomoću alata Minimap2 koristimo kao ulazne točke programa.

Ulazni podaci:

- preklapanja između contig-a i očitavanja u PAF formatu dobivena korištenjem alata Minimap2 nad datotekom sa skupom contig-a i datotekom sa skupom očitavanja
- međusobna preklapanja očitavanja u PAF formatu dobivena korištenjem alata Minimap2 nad dvije iste datoteke sa skupom očitavanja

Izlazni podaci:

- poboljšani skup sastavljenih contiga u FASTA formatu

Sada kada imamo sve potrebne podatke možemo konstruirati graf preklapanja. Graf preklapanja sastoji se od dvije vrste čvorova: usidreni čvorovi (*anchoring nodes*) koji predstavljaju unaprijed sastavljane contig-e i čvorovi očitavanja (*read nodes*). Veza između dva čvorova predstavlja preklapanje tih dvaju čvorova. Kada je graf konstruiran, slijedi traženje optimalnih staza iz svakog usidrenog čvora do skupa mogućih završnih usidrenih čvorova.

3. Zaključak

Zaključak.