

Zadatak 12

Implementirati na programskom jeziku Python:

1. Algoritam za indeksirano pretraživanje stringova u zadatom tekstu koristeći Burrows-Wheeler transformaciju i FM index. Inicijalna verzija algoritma treba da bude realizovana na tradicionalan način opisan na predavanju, bez optimizacije memorije i vremena izvršavanja (**10 poena**).
2. Algoritam za globalno poravnavanje stringova (global alignment - Needleman-Wunsch) koji koristi scoring tabelu (scoring matrix) kao ulazni parametar komandne linije programa (**5 poena**).
3. Program prima kao ulaz referencu ([FASTA](#) fajl) i kolekciju ridova ([FASTQ](#) fajl) kao argumente komandne linije i za svaki od ridova vraća listu torki (pozicija gde se rid mapira na referencu, ocena optimalnog poravnanja na toj poziciji i edit transkript) - liste treba sortirati po oceni od najveće ka najmanjoj. Koristiti "Seed & Extend" metod gde se seed uzima sa početka rida (dužina seed-a je ulazni parametar), za lociranje pozicije seed-a u referenci koristiti algoritam iz tačke (1). Zatim, preostali deo rida treba poravnati na odgovarajući deo reference iza lociranog seed-a koristeći algoritam iz tačke (2). Deo reference na koji se preostali deo rida poravnava treba biti duži od preostalog dela rida za 0 - 3 nukleotide (ulazni parametar "margin"). Za svaki od ridova izvršiti pretraživanje i za njegov reverse komplement, jer je moguće da potiče sa backward strenda referentnog genoma. (**5 poena**).
4. Uzeti za dužinu seed-a 10 nukleotida, "margin" parametar 2, vrednost poklapanja (match) [0, 1, 2], mutacije (mismatch) [-3, -2] i insercije ili delecije (gap) [-7, -5]. Dati dijagrame poređenja dobijenih najboljih alajnmenta za svaki rid u odnosu na rezultate dobijene [BWA-MEM](#) alatom. Pozicije najboljih alajnmenta BWA-MEM-a biće dostupne u njegovom izlaznom SAM fajlu. Testirati za svih 12 tabela dobijenih pomenutim kombinacijama parametara. Za testiranje koristiti dati [referentni genom](#) i [kolekciju ridova](#). Prikazati rezultate tabelarno i grafički (**5 poena**).

Pripremiti prezentaciju (Google slides ili Power point) implementiranog rešenja, kao i samih rezultata (**5 poena**).

Pripremiti video prezentaciju projekta (5 - 10 minuta trajanja) koja će biti dostupna na YouTube ili drugom on-line video servisu (**10 poena**).