

Zadaci sa vezbi:

1. a) Ucitati FASTA _____ koristeći PYSAM biblioteku i izdvojiti sekvencu na ___ chr od ____:____ pozicije u promenljivoj 'sekvenca'.
b) Napisati funkciju kamerizuj(sekvencu, k) koja za argumente ima DNK sekvencu i duzinu k-mera, a vraća listu svih jedinstvenih k-mera prisutnih u sekvenci i pozvati je na izdvojenoj sekvenci. Koliko je jedinstvenih k-mera dužine 4?
c) Napisati funkciju debruijnizuj(read, k) koja za DNK sekvencu i duzinu k-mera, za svaki jedinstveni k-mer u sekvenci vraća listu njegovog levog k-1-mera i njegovog desnog k-1-mera u paru.
Npr: print(debruijnizuj("ACGCGTCG", 3))
[(('AC', 'CG'), ('CG', 'GC'), ('GC', 'CG'), ('CG', 'GT'), ('GT', 'TC'), ('TC', 'CG'))]
d) Koju vrednost vraća len(debruijnizuj(sekvencu, 3))?
2. a) Ucitati _____ FASTA file koristeći PYSAM python biblioteku.
b) Odstampati regione svih kontiga u FASTI. Koji je njihov broj?
c) Koja je dužina hromozoma 17?
d) Dohvatiti region 43044295:43125370 sa hromozoma 17 i odrediti procenat G i C baza.
e) Dohvatiti region 50100:50200 sa hromozoma 1 i odrediti koliko kojih baza je prisutno. Koliko A,T,C,G?