Program za pronalaženje restrikcionih mesta

Restrikcioni enzim je enzim koji ima sposobnost da prepozna kraću sekvencu DNK (uglavnom dužine 4-8 nukleotida), koja se naziva restrikciono mesto, i da potom seče DNK ili u okviru tog restrikcionog mesta ili uzvodno ili nizvodno.

Poželjno je da restrikciono mesto bude palindromsko – ukoliko ga prevedemo u komplementaran lanac od nazad će biti čitano isto kao restrikciono mesto koje smo dobili. S obzirom da se DNK sastoji od dva lanca i da se njihovi nukleotidi uvek sparuju Adenin-Timin i Citozin-Guanin, prevođenjem restrikcionog mesta kao što je GATC u komplementarni lanac dobijamo CTAG. Dva lanca DNK čitaju se iz suprotnih smerova: 5’-3’ nazivamo lanac koji čitamo sa leva na desno i on se često zove glavni lanac (u slučajevima kada dobijamo samo jedan lanac, kao GATC ili neka jednolančana sekvenca, uglavnom se podrazumeva da je u pitanju 5’-3’ lanac), dok se njemu komplementarni 3’-5’ lanac čita sa desna na levo. Zato, ukoliko se sekvenca komplementarna restrikcionom mestu čita isto od nazad, onda znamo da će enzim prepoznati isto mesto na oba lanca DNK.

Restrikciona mesta unutar sebe nekada mogu sadržati varijabilne pozicije: GANNTC je restrikciono mesto koje u sredini može sadržati bilo koja dva nukleotida, dok god počinje sa nukleotidima GA i završava se sa TC.

Restrikcioni enzimi se mogu naći kod bakterija i archea gde predstavljaju odbrambeni mehanizam (lako isecanje i izbacivanje strane npr. virusne DNK) i dosada ih je otkriveno preko 3 000. Preko 600 od tih restrikcionih enzima dostupno je komercijalno (a sve je rasprostranjenije i pravljenje veštačkih) i predstavljaju ključni alat u DNK modifikaciji, analizi nepoznate DNK i kloniranju.

Fajl **chkgen.py** sadrži prvi deo programa gde proveravamo da li je sekvenca DNK genomska. Sekvenca je genomska ukoliko sadrži samo slova G, A, T i/ili C.

Fajl **cutgen.py** je deo programa koji se odnosi na restrikciona mesta.

Mesto sečenja restrikcionog enzima označava se sa ’’^’’. Ukoliko se mesto sečenja nalazi uzvodno ili nizvodno, ispred ili iza restrikcionog mesta biće onoliko ‘’-‘’ koliko nukleotida ispred, odnosno iza, seče restrikcioni enzim. Bilo je potrebno definisati funkciju (iskasapi) koja će vratiti poziciju karota unutar restrikcionog mesta (kao I da li se nalazi ispred ili iza I koliko pozicija je u pitanju) I restrikciono mesto bez označenog mesta sečenja.

U restrikcionom mestu možemo imati varijabilne pozicije: N (G, A, T ili C), M (A ili C), K (G ili T), R (A ili G), Y (C ili T). Funkcija dodatni\_bp omogućava proveravanje da li je restrikciono mesto palindromsko nakon što smo restrikciono mesto pretvorili u komplementarnu sekvencu (odnosno niz) i okrenuli ga (da bismo čitali od nazad). Ukoliko restrikciono mesto nije palindromsko program će odštampati upozorenje ’’Restrikciono mesto nije palindromsko!’’.

Jednom while petljom pozivamo funkciju poredjenje, koja poredi DNK sekvencu i restrikciono mesto kako bi pronašla gde se u DNK sekvenci nalazi to restrikciono mesto, i funkciju secenje, koja sekvencu DNK seče u fragmente (i ukoliko enzim seče ispred restrikcionog mesta omogućava nastavak poređenja sekvence i restrikcionog mesta posle mesta sečenja, a ne posle pronađenog restrikcionog mesta). Funkcija secenje takođe sadrži upozorenje ’’Prepoznalo se na mestu, nije se seklo’’ ukoliko je restrikciono mesto prepoznato na samom početku, odnosno kraju, DNK sekvence koje seče nekoliko mesta ispred, odnosno iza, restrikcionog mesta.

Program će odštampati leva\_strana (što odgovara odštampanom ’’mesto sečenja), sekvenca[leva\_strana:] (što odgovara odštampanom ’’fragment’’) i len(sekvenca)-leva\_strana (što odgovara odštampanom ’’dužina fragmenta’’).

Fajl **proteingen.py** odnosi se na pretvaranje DNK fragmenata koje smo dobili u proteinske sekvence. DNK sekvenca se u proteinsku sekvencu pretvara redom uzimanjem po tri nukleotida koja pretvaramo u amino kiselinu koja im odgovara. Ukoliko dužina fragmenta nije deljiva sa 3, nije moguće ceo fragment pretvoriti u proteinsku sekvencu i u tom slučaju program će odštampati ostatak.

Fajl **sequence.txt** je fajl koji sadrži sekvencu DNK koju program učitava.

Fajl **aafrgen.sh** je bash script koji povezuje tri programa – omogućava učitavanje rezultata prvog python programa u drugi (i zaustavljanje ukoliko data sekvenca nije genomska), kao i učitavanje rezultata drugog python programa (fragmenti) u treći.